

**A CASE STUDY OF THE MANAGEABILITY AND
UTILITY OF ASSESSMENT IN THREE
NEW ZEALAND PRIMARY SCHOOLS
1993-2006**

by

JOHN RICHARD YOUNG

**A thesis submitted to Victoria University of Wellington in fulfilment
of the degree of Doctor of Philosophy in Education**

**Victoria University of Wellington
2009**

ABSTRACT

This study is concerned with the manageability of assessment in New Zealand schools. In order for assessment to be effective in schools, it needs to be manageable. This is especially important for New Zealand schools in the light of the changes in administration and curriculum introduced in the early 1990s. These reforms which were termed *Tomorrow's Schools* introduced radical changes to New Zealand schools in the areas of governance, management, the curriculum, assessment, and quality assurance.

In the area of assessment, primary schools were presented with a number of challenges which included a requirement to assess a highly structured curriculum with close to 1000 achievement objectives, many of which were open to different interpretation. There was also an expectation to assess against the structure of this new rapidly introduced curriculum, without any tools provided for such assessment. The New Zealand system also departed from other countries in that it made no distinction in terms of importance in teaching and assessment between what are usually considered the core activities of primary school education, literacy and numeracy, and other learning areas. In light of these challenges it is not surprising that a number of the assessment processes that schools developed and used were questionable in terms of manageability and utility.

The aim of this research is to examine the effect of government policy and international movements in assessment on the manageability of assessment practice in New Zealand primary schools during the period 1990-2006. The research also examines what is meant by the terms *manageability* and *utility* when applied to the context of primary school assessment.

By means of case study research, the investigation explores assessment practice in three Wellington primary schools during 2006-2007. The case studies sought to understand current assessment practices in terms of the reforms of the early nineties, as well as more recent developments in assessment in New Zealand.

The research illustrated that while these three New Zealand schools still faced some major issues in terms of the manageability and utility of assessment, their views were more optimistic and positive than earlier research studies had reported.

I.

ACKNOWLEDGEMENTS

There are a number of people whose support I would like to acknowledge in the completion of this thesis. Firstly to my supervisor Cedric Hall for his patience, guidance and wisdom over the last 5 years.

Thanks also need to be expressed to the staff and students of St Joseph's School Upper Hutt. Many of the classes of the school had to experience large quantities of assessment as I sought to understand many of the new assessment tools. I would also like to express my gratitude to the BOT of St Joseph's for the support which allowed me to undertake this study and their granting of leave to me during two terms in 2007 to work on this research. I am deeply grateful to New Zealand Ministry of Education's Teacher's Study Award Scheme that granted me 20 weeks paid leave in 2007. Without this scheme it would have been extremely difficult to complete this project.

I would also like to express my deep thanks to Radha Hendriks, Pamela Knee, Fiona Bolger, Therese Young, Anna Young and Bill Sutton for their editorial help.

It is important also that I acknowledge the support of a group of Wellington principals known collectively as "McGill's" after our first, and favourite first Friday of the month watering hole. To Bill Turley, Bill Sutton, Paul Roche, Dennis Thompson, Mike McGimpsey, Alan Fleming, Neil Worboys, Peter Ahern, John Masson, Neil Whittington, and the two Kevin Ryans, I would like to thank you for your support and your patience in listening to, and responding to my thoughts on assessment over the last five years.

Finally I would like express my thanks to my family: Anna, Therese, Rebecca, Daniel and Claire Young for their support and encouragement. In particular my deep thanks go to my wife Anna and her IT and layout skills without which this thesis would not have been completed.

ABBREVIATIONS AND ACRONYMS

Abbreviation

ABeL	Assessment for Better Learning
ARB	Assessment Resource Bank
asTTle	Assessment Tools for Teaching and Learning
ATOL	Assessment to Learn
BSM	Beginning School Mathematics
BOTs	Boards of Trustees
CRA	Criterion Referenced Assessment
ERO	Education Review Office
IEA	International Association for the Evaluation of Education
LPDP	Literacy Project Professional Development
MoE	Ministry of Education
MinNZC	Mathematics in the New Zealand Curriculum
NAGs	National Administration Guidelines
NCEA	National Certificate in Educational Achievement
NEGs	National Education Goals
NEMP	National Education Monitoring Project
NZCER	New Zealand Council for Educational Research
NZCF	New Zealand Curriculum Framework

NZEI	New Zealand Education Institute
NZPF	New Zealand Principals' Federation
NZQA	New Zealand Qualifications Authority
OBE	Outcomes Based Education
P & A	Progress and Achievement Register
PIRLS	Progress in International Reading Literacy Study
PAT	Progressive Achievement Test
PISA	Programme for International Student Assessment
SBA	Standards Based Assessment
SEA	School Entry Assessment
SATs	Standardised Assessment Tasks
STAR	Supplementary Test of Reading Achievement
TIMSS	The Third International Mathematics and Science Study
TOSCA	Test of Scholastic Abilities
The Exemplars	The New Zealand Curriculum Exemplars

Glossary

Cumulative cards

A cumulative card is a card that records student details and a summary of their achievement. Their use is not compulsory, and while previously very common in schools use is now declining.

Cluster

A cluster is a term used in New Zealand to describe a grouping of schools in a geographical area. The organisation of clusters has no official status but is usually a basis for forming school groups linked by membership of a local principals' association. In the last ten years it has been increasingly common for clusters to undertake some professional development together and administer special needs resourcing.

Professional Development Contract /LPDP/ATOL

A Professional Development (PD) Contract is a process of professional development in a school that is usually of between one or two years duration. It is a whole school development involving all teaching staff. PD contracts are commonly led by an outside facilitator. A school undertaking a PD contract is usually required to make some formal commitment to implementing the content of the contract. Literacy Professional Development Contract (LPDP) is a PD contract focusing on either reading or writing. An Assessment to Learn (ATOL) formerly ABeL is an assessment PD contract.

Scale A Teacher

A scale A teacher is someone who is responsible for their own class teaching but has no designated management responsibilities in a school.

Syndicate

A syndicate is the term used in most New Zealand schools to describe a grouping of classes at a similar level in a school. Syndicates usually include three to four classes.

Syndicate leader

A syndicate leader is the person who has management responsibility for a syndicate.

Topic

Topic is the term used in many New Zealand primary schools to describe teaching in social studies, science, technology and health. It can be common for these subjects to be integrated with the English curriculum. It is very common for Topic to be taught in the afternoon in blocks of four to six weeks on one area of the curriculum.

TABLE OF CONTENTS

ABSTRACT	i
ACKNOWLEDGEMENTS	iii
ABBREVIATIONS AND ACRONYMS	iv
TABLE OF CONTENTS	viii
LIST OF TABLES	xiii
LIST OF FIGURES	xiii

CHAPTER ONE1

1.1 Introduction to Chapter One	1
1.2 The Rationale for the Research.....	1
1.3 The Context of the Study	2
1.3.1 The New Zealand Education Reforms	2
1.3.2 Recent Developments in New Zealand Assessment Practice	4
1.3.3 New Zealand Assessment Literature	5
1.4 Research Focus	7
1.4.1 Research Context	7
1.4.2 Research Questions	7
1.5 Methodology	8
1.5.1 Methodology	8
1.5.2 Sample.....	8
1.5.3 Case Study Research.....	9
1.6 Structure of the Thesis	10
1.7 The Researcher's Perspective	11

CHAPTER TWO: THE CONCEPTS OF MANAGEABILITY AND UTILITY IN PRIMARY SCHOOL ASSESSMENT16

2.1 Introduction to Chapter Two	16
2.2 The Manageability of Assessment.....	17
2.2.1 Manageability in the General Assessment Literature	17
2.2.2 Manageability and Utility	18
2.3 Components of the Manageability of Assessment	19
2.4 Manageability of Assessment Systems.....	22
2.5 Four Contexts relating to the Manageability and Utility of School Assessment.....	23
2.5.1 The New Zealand Context of Prescribed and School-Based Assessment ..	25
2.6 Manageability in Context: the UK Experience.....	25
2.7 Manageability is about People.....	26
2.8 The Purpose of Primary School Assessment	30
2.9 The Relationship between Purpose, Validity, Manageability and the Utility of Assessment.....	33
2.10 Definitions of Manageability and Utility.....	36

2.11 Summary Chapter Two	39
--------------------------------	----

CHAPTER THREE: ASSESSMENT IN NEW ZEALAND PRIMARY SCHOOLS 1990-200640

3.1 Introduction to Chapter Three	40
3.2 Assessment Activity in New Zealand Primary Schools in the 1980s	41
3.3 The Accountability Movement and Outcomes-Based Education	44
3.3.1 The Development of the Accountability Movement and OBE.....	45
3.3.2. The Nature of Outcomes Based Education.....	46
3.3.3 The Curriculum and Assessment in Outcomes Based Education.....	47
3.3.4 Outcomes Based Education, Standards Based Assessment, and Criterion- Referenced Assessment	48
3.3.5 Implementation in New Zealand.....	49
3.3.6 Criticism of OBE in New Zealand.....	49
3.4 Four Major Assessment Challenges	51
3.5 The Lack of Direction Given to Schools in Assessment.....	52
3.6 The Structure of the Curriculum.....	56
3.6.1 asTTle and the Levels of the Curriculum	62
3.7 The Lack of Distinction between the Core Curriculum and the Other Areas of Learning	64
3.8 The Impact of the Education Review Office	66
3.8.1 ERO and National Testing	69
3.8.2 The Changing Role of ERO	70
3.9 The Response of Schools to the Assessment Requirements of the Reforms.....	70
3.10 The Professional Development of Teachers: Assessment for Better Learning	72
3.11 Planning and Reporting	72
3.12 Recent Developments	73
3.13 International Influences	74
3.14 Authentic or Performance Assessment.....	75
3.15 Formative Assessment	83
3.16 Evidence-Based Teaching.....	89
3.17 International and National Testing	92
3.18 Summary Chapter Three.....	97

CHAPTER FOUR: ASSESSMENT TOOLS AND PROCEDURES USED IN NEW ZEALAND PRIMARY SCHOOLS100

4.1 Introduction to New Zealand Assessment Tools	100
4.2 Progressive Achievement Tests (PATs)	100
4.3 Running Records.....	103
4.4 National Education Monitoring Project (NEMP)	107
4.5 The Assessment Resource Banks (ARB)	108

4.6 School Entry Assessment.....	111
4.7 STAR	113
4.8 The New Zealand Curriculum Exemplars	115
4.9 Numeracy Project	117
4.10 asTTle (Assessment Tools for Teaching and Learning).....	118
4.10.1 Development of asTTle.....	118
4.10.2 Development of asTTle and Technical Information	120
4.10.3 The Manageability of asTTle.....	123
4.11 Recent Developments	123
4.12 Mathematics Assessment.....	124
4.13 Assessment in the Non-Core Curriculum Areas: the Visual Arts ..	126
4.14 Summary of New Zealand Assessment Tools.....	129
CHAPTER FIVE: METHODOLOGY.....	131
5.1 Introduction to Methodology	131
5.2 A Constructivist World View	131
5.3 Approaches to Research.....	132
5.4 Positivistic Paradigm	133
5.5 The Interpretivist Paradigm and Qualitative Research.....	135
5.6 Pragmatism - Mixed Methods Approaches	137
5.7 Selection of Methodology: Case Study	138
5.8 Semi-structured Interviews	141
5.9 Document Examination	142
5.10 Focus Groups	142
5.11 Conducting the Research	143
5.11.1 Research Sample	143
5.11.2 Selection of Case Study Schools.....	144
5.11.3 Recruitment of Schools.....	145
5.11.4 Interview Procedures	145
5.11.5 Documentation Examination in the Case Study Schools.....	146
5.12 Data Analysis	146
5.13 Follow-up Activities	147
5.14 Challenges.....	148
5.15 Trustworthiness, Validity, Reliability and Transferability.....	148
5.15.1 Perspective of the Researcher	148
5.15.2 Validity in Research.....	149
5.15.3 Triangulation.....	150
5.15.4 Member Checks	152
5.15.5 Participatory Models of Research	153
5.15.6 Researcher Bias.....	153
5.15.7 Reliability.....	153
5.15.8 Transferability	154
5.16 Ethical Issues	155
5.17 Summary of Methodology	156

CHAPTER SIX: CASE STUDY SCHOOLS157

6.1 Introduction to Chapter Six.....	157
6.1.1 Research Process.....	157
6.1.2 Structure of the Case Study Report.....	158
6.2 Case Study One: Endeavour School	159
6.2.1 Description of Endeavour School	159
6.2.2 Assessment Prior to and in the Early Years of the Reforms	162
6.2.3 Endeavour Assessment Practices, February 2007	162
6.2.4 Assessment Goals	163
6.2.5 The Implementation of Assessment.....	168
6.2.6 Challenges in Implementation	172
6.2.7 The Manageability and Utility of Individual Assessment Tools	173
6.2.8 Assessment Software	177
6.2.9 Formative Assessment	178
6.2.10 Summative Assessment	178
6.2.11 Summary of Assessment Challenges, Manageability and Utility.....	179
6.2.12 Summary of Assessment at Endeavour School	180
6.3 Case Study Two: Tasman School	182
6.3.1 Description of Tasman School.....	182
6.3.2 Assessment Practices, February 2007	184
6.3.3 Seven Major Themes in Assessment at Tasman School.....	185
6.3.4 Assessment Tools and Procedures	189
6.3.5 The Manageability and Utility of Assessment.....	193
6.3.6 ERO Reviews.....	193
6.3.7 Summary of Assessment at Tasman School	199
6.4 Case Study Three: Kupe School	201
6.4.1 Description of Kupe School.....	201
6.4.2 Assessment Themes	203
6.4.3 Individual Tools and Procedures at Kupe School.....	209
6.4.4 Assessment issues for Kupe.....	213
6.4.5 Summary of Assessment at Kupe School	214

CHAPTER SEVEN: SUMMARY AND INTEGRATION.....216

7.1 Introduction.....	216
7.2 What were the Major Influences and Challenges to Assessment in New Zealand Primary Schools since 1990?	217
7.3 Influences and Challenges in the Three Case Study Schools.....	221
7.3.1 Structure of the Curriculum/Achievement Objectives.....	221
7.3.2 Portfolios and Student Folders.....	222
7.3.3 The Assessment of the Non-Core Curriculum.....	222
7.4 Three Waves of Assessment Practice	224
7.5 Factors that Influence the Establishment of Assessment and Reporting Systems	228
7.5.1 School Leaders.....	228
7.5.2 Professional Development	231
7.5.3 Networks with Other Schools	233

7.5.4 The Influence of ERO	234
7.5.5 Ministry of Education	235
7.6 Four Phases of Assessment Change	236
7.7 Manageability and Utility in the Case Study Schools	240
7.8 Manageability and Utility in New Zealand Primary Education	242
7.9 Recent Developments – Research Question Three.....	246
7.9.1 Formative Assessment	246
7.9.2 ARBs, The Exemplars, NEMP tasks	247
7.9.3 School Entry Assessment.....	248
7.9.4 asTTle	248
7.9.5 STAR	249
7.9.6 Numeracy Project Assessment.....	249
7.9.7 Planning and Reporting Requirements	250
7.9.8 Summary of Recent Developments – Research Question Three	250
 CHAPTER EIGHT: LIMITATIONS OF THIS RESEARCH, RECOMMENDATIONS AND FURTHER RESEARCH.....	 252
8.1 Limitations of this Research	252
8.2 Recommendations.....	254
8.2.1 Non-Core Curriculum	254
8.2.2 The Written Recording of Formative Assessment Feedback Comments	255
8.2.3 The Frequency of Assessment	255
8.2.4 The Economics of Assessment Development in New Zealand Schools.....	256
8.3 Suggestions for Further Research.....	256
8.3.1 On-going Research on Assessment Issues in New Zealand Schools.....	256
REFERENCES	258
APPENDIX A: Principals and Senior Management Interview Form	279
APPENDIX B: Teacher Interviews	284
APPENDIX C: Consent Form - Principal	288
APPENDIX D: Consent Form - Staff Member	289
APPENDIX E: Letter to Principals/Staff Member.....	290
APPENDIX F: Focus Group Process	291

LIST OF TABLES

Table 2.1: Components of manageability and utility emphasised in psychometric tests	21
Table 3.1: New Zealand International Assessment Results	93
Table 3.2: Assessment in New Zealand primary schools 1989-2006: some relevant landmarks	98
Table 4.2: Mathematics Assessment Tools in use in New Zealand Schools	125
Table 5.1: Common elements in data from within individual case study schools.....	151
Table 5.2: Common elements in data across all three schools	152
Table 6.1: Description of Endeavour School	160
Table 6.2: Assessment and Reporting Practices at Endeavour School as at February 2007.....	161
Table 6.3: Tasman School Description (2005 ERO Report)	183
Table 6.4: Assessment and Reporting at Tasman School February 2007.....	184
Table 6.5: Kupe School Description	202
Table 6.6: Assessment Practices as at Kupe School as at December 2006	202
Table 7.1: Assessment in New Zealand primary schools 1989-2006.....	218

LIST OF FIGURES

Figure 3.1: Curriculum Levels and Year Levels in the New Zealand Curriculum.....	63
Figure 4.1: Levels of the New Zealand Curriculum and Year Levels	122
Figure 7.1: Three Waves of Assessment Practice in New Zealand Primary Schools 1989-2006	227
Figure 7.2: Four Phases of Assessment Change in New Zealand Primary Schools..	237
Figure 7.3 The relationship between manageability, utility, reliability and validity	244

CHAPTER ONE

1.1 Introduction to Chapter One

In order for assessment to be effective in schools, it needs to be manageable. This is especially important for New Zealand schools, in the light of the changes in administration and curriculum introduced in the early 1990s. As will be argued in this thesis, these changes have meant that assessment expectations for primary schools have been often unrealistic and lacking in appropriate guidance and direction.

While there are many important indicators of quality in assessment practice such as reliability, validity and the meeting of ethical standards, manageability and utility are also key features of successful assessment. As will be argued, manageability, and to a lesser extent, utility, have often been neglected in the development of assessment expectations, tools and practices in New Zealand schools, giving rise to a number of implementation problems.

1.2 The Rationale for the Research

The rationale for this research is based on three related themes. The first theme concerns a desire to understand the challenges New Zealand schools faced arising from the structural and curriculum reforms that were introduced during the period 1989 to 1993. A major outcome of these reforms was the increased assessment requirements that schools faced (Fancy, 2004a).

The requirement of increased assessment associated with the reforms led to questions being asked about the manageability and utility of a number of the assessment activities schools undertook during the mid to late nineties and the pressure they were placing on teachers and principals (Livingstone, 1994; MoE, 1999b; Renwick & Gray, 1995; Thrupp, Harold, Mansell, & Hawksworth, 2000; and Wylie 1997, 1999). However, in order to answer questions regarding the manageability and utility of assessment activities it is essential to develop an understanding of what is meant by these two concepts. For this reason an attempt is made to develop an understanding of the concepts of the manageability and utility of assessment in the context of New Zealand primary education. Such an understanding has the potential to improve the implementation of assessment in schools. For this reason the development of an

understanding of the concepts of the manageability and utility of assessment is the second major rationale for the research.

A particular feature of the first years of New Zealand reforms was that schools were asked to assess a curriculum without the provision of assessment tools to carry out this process (ERO, 1995). However, since the new millennium, New Zealand primary schools have been supplied with a rapidly increasing number of assessment tools. These include assessment activities associated with the Numeracy Project, the Supplementary Test of Reading Achievement (STAR) (Elley, 2001), the Exemplars (Chamberlain, 2001; MoE, 2003a), the revised PAT tests (Darr, Neill & Stephanou, 2006; Darr, McDowall, Ferral, Twist & Watson, 2008) and asTTle (MoE, 2007). Attempting to understand how New Zealand primary schools have adapted to the introduction of these tools is the third rationale for this study.

In summary, the rationale for this research is an attempt to understand the impact of the structural and curriculum reforms on the manageability and utility of assessment activity carried out in New Zealand schools during the period 1993 to 2006. It is believed that greater understanding of these concepts could lead to more useful and manageable assessment practices in the future.

1.3 The Context of the Study

1.3.1 The New Zealand Education Reforms

The reforms that began in New Zealand in 1989 (Lange, 1988) were arguably the most fundamental changes in the country's education since the introduction of compulsory education in 1877. Schools were to become self governing and self managing with the establishment of Boards of Trustees (BOTs).

Self management meant that the BOTs were responsible for property, financial operations, employment of all staff, as well as developing their own individual sets of policies and procedures. The role of the principal expanded from the traditional focus on such elements as the curriculum, student behaviour and the leading of teachers, to becoming property managers, and developing an understanding of both accounting principles and employment law. Greater accountability was introduced into the system with the former inspectorate, which had a strong element of advice and guidance,

being replaced by an independent audit agency, the Education Review Office (ERO). ERO reports were to become public documents and results were published in the local press.

While self management gave schools more local control on some aspects of decision making, the curriculum became more centrally controlled and directive. The New Zealand Curriculum Framework (NZCF) (MoE, 1993a) was published in 1993 and was followed by the roll out of the various curriculum documents. These seven new curriculum statements were introduced at a very rapid pace: Mathematics (MoE, 1992) (which actually preceded the NZCF); Science (MoE, 1993b); English (MoE, 1994a); Technology (MoE, 1995); Social Studies (MoE, 1997); Health and Physical Wellbeing (MoE, 1999a); and The Arts (MoE, 2000a). The rapid pace of introduction put a great deal of pressure on schools to develop systems to implement these radically new approaches (Fancy, 2004a). The curriculum statements were organised under eight vertical levels and a number of strands. Each strand of the curriculum was broken down into achievement objectives (over 200 in the *Mathematics in the New Zealand Curriculum* from level one to five alone).

Willis (1992) and Elley (1993) highlighted how the New Zealand reforms were based on the earlier English reforms. While there were many similarities in terms of the philosophy of competition, increased accountability and the structuring of the curriculum around a series of levels, the approach to assessment differed quite markedly. The English had a far more directed structure with the prescribed system of Standardised Assessment Tasks (SATs). In New Zealand, schools were expected to assess students but were given little direction on how to carry this out (Capper, 1991). Such decisions were to be made at the school level. The lack of direction given to New Zealand schools also produced significant, if different challenges, as documented by ERO in 1995. These included concerns over the large number of achievement objectives to be assessed, the lack of tools to assess the new curriculum, and the inefficiency of all 2700 New Zealand schools developing their own assessment systems.

The MoE's National Education Goals (NEGs), National Administration Guidelines (NAGs) and National Achievement Objectives which took effect in April 1993 spelt out the requirements of BOTs. In terms of assessment the NAGs made no distinction

between the traditional core areas of literacy and numeracy, and other subjects. Schools were expected to assess in detail all areas of the curriculum.

1.3.2 Recent Developments in New Zealand Assessment Practice

Since the reforms a number of assessment tools and developments have been introduced. The first of these was the National Education Monitoring Project (NEMP) (Flockton, 1995). The Assessment Resource Banks (ARBs) developed by the New Zealand Council for Educational Research (NZCER) were the first tools introduced to assess the new curriculum and were made available to schools in 1996. The Assessment for Better Learning (ABeL) professional development contract began in 1993. The title later changed to Assessment to Learn (ATOL). These programmes have aimed to develop teacher knowledge of effective assessment and to raise students' levels of achievement.

In the late 1990s there was a push for national testing led by ERO (1998) and supported by the National Government Ministers of Education, Wyatt Creech and Nick Smith (MoE, 1998).

At the same time concern regarding the reading achievement of many New Zealand students led to the establishment of the Literacy Task Force (MoE, 1999b) with the subsequent development of the Supplementary Test of Reading Achievement (STAR) in 2001 by NZCER (Elley, 2001).

From the late 1990s the Numeracy Project (Higgins, 2002) was introduced progressively into schools beginning with those in the lowest deciles. This method of teaching numeracy included a comprehensive one-to-one assessment which was radically different from previous approaches to mathematics assessment, especially in the middle and senior school which had formerly been largely based on school developed class tests.

The new millennium began with the election of a Labour Government and the appointment of Trevor Mallard as Minister of Education. The pace of change in the development of assessment tools was accelerated. While the plan for national testing was abandoned, the development of the actual tests, which were titled Assessment Tools for Teaching and Learning (asTTle) (MoE, 2007), was continued with their use

being voluntary. The first version of asTTle became available in 2003. The New Zealand Curriculum Exemplars (the Exemplars) (Chamberlain, 2001; MoE, 2003a), a set of assessment and teaching tools based on the various curriculum statements, were also developed by the MoE.

Under Mallard's ministry, professional development was increased with a greater number of schools undertaking the revised ATOL contract with its emphasis on formative assessment. Also a large number of schools undertook Literacy Professional Development Project (LPDP) contracts for professional development which also had a strong emphasis on formative assessment. In a new development, and contrasting with the early philosophy of competition, this period also saw schools working together in clusters (Fancy, 2004a).

In 2001 the government introduced the *Education Standards* (MoE, 2001d) legislation which required schools to set annual targets and report on progress to their communities and the Ministry. Five years later in 2006 the Ministry notified all schools that over the next two years they would be required to purchase and use an approved school management software package. All these programmes included a large element of assessment recording and reporting. The most recent development was the publication of the revised draft national curriculum (MoE, 2007a) released to schools in July 2006. The implications of this for assessment have yet to be worked through.

1.3.3 New Zealand Assessment Literature

A brief background on New Zealand assessment literature over the last 20 years provides further context for the position of this research. A review of the various perspectives of writers on the subject of assessment in New Zealand primary schools since 1990 shows that there has been little empirical work that looks specifically at actual classroom and school practice. The little work that has taken place has not been recent enough to take account of the rapidly changing assessment landscape. Assessment and its relationship to the curriculum have been examined in the context of changes in the New Zealand education system in the period 1990 to the present, by a number of writers. Some of this research (O'Neil, Clark, & Openshaw, 2004; Willis, 1992) focused more on the ideological perspective of curriculum and assessment

reform than on the effects on practice and learning. However, Elley (2004) and Thrupp, Harold, Mansell, and Hawksworth (2000) examined the reforms from the perspective of both ideology and practice, while Livingstone (1994), Renwick and Gray (1995) and Wylie (1997, 1999) investigated the impact of the reforms in general on teachers, principals and schools and each included substantial sections on assessment. Livingstone conducted research into the workloads of 157 Wellington teachers and identified the development of new assessment procedures as one of the five most stressful workload factors. Renwick and Gray reported that teachers had found the pace of the curriculum reforms challenging with most reporting increased workloads. Assessment was an issue with a large number of teachers. There was, in Renwick and Gray's words, "uncertainty about how much should be assessed and appropriate assessment methods" (p. 9). Wylie in her conclusion from a series of teacher surveys up to 1997 reported that teachers had greatly increased the amount of assessment they were doing.

Croft, Stafford, and Mappa (2000) and Dunn, Stafford and Marston (2003) examined the extent of use of individual assessment tools in New Zealand schools and teachers' views on their effectiveness. However, there are few actual studies of assessment in New Zealand conducted in the last ten years, although the works of Brown (2002), Dixon (1999), Hill (2000) and Knight (2000) stand out as examples of studies where assessment was the primary focus, and in Hill's case, the research was based on actual classroom practice. Knight brought to her writing on assessment for accountability the perspective of a school facilitator who was working with a number of schools in the area of assessment.

Through questionnaires and interviews with teachers in four Auckland schools, Dixon (1999) illustrated the gap between policy and practice in assessment. Her research identified the impact of the Education Review Office on assessment practice in New Zealand schools in the late nineties. The study also illustrated the growing understanding of formative assessment by teachers.

In comparison to the research of Dixon (1999) and Hill (2000), this thesis includes the most recent developments in assessment, but the most significant difference is that this study focuses more on the school as a system and has a heavy emphasis on the role of principals and other school leaders in assessment activity, whereas Hill's work

focused on teachers in the classroom, Brown's (2002) on individual teachers' perceptions of assessment in general, and Dixon excluded principals from her study. Furthermore, this work looks at teachers' views on specific individual assessment tools and procedures. In other words, a key distinguishing feature of this research is that the school as a unit of study is the focus through an examination of the views of both teachers and school leaders, as well as an analysis of the interaction between teachers and leaders, and between the school and its community.

1.4 Research Focus

1.4.1 Research Context

Given the rationale and context given above, this research is focused on how assessment activity is implemented in schools. It examines how schools make sense of the requirements and attempt to establish assessment and reporting systems. Central to this examination are explorations of how manageable and useful are the assessment activities schools undertake.

1.4.2 Research Questions

The above aim forms three research questions:

1. What were the major influences and challenges to assessment in New Zealand primary schools since 1990?
2. How did three New Zealand primary schools establish assessment and reporting systems during the period that were manageable and maintained a focus on learning?
3. To what extent do the staff of three case study schools find the most recent changes in assessment practice/tools manageable and provide useful information?

1.5 Methodology

1.5.1 Methodology

This section briefly outlines the methodological approach used in this study. This will be covered in greater detail in Chapter Five. The research was based on a review of the assessment literature (question one) and case studies of three schools, examining how they established and developed assessment systems from the introduction of the Tomorrows Schools reforms to 2006 (questions two and three). The approach included semi-structured interviews with principals, other school leaders, and teachers. It also included document examination of such artefacts as school assessment policies, strategic plans, BOT reports, ERO reports, and reporting to parents. The research also examined how the three schools were implementing the most recent changes in the assessment landscape.

1.5.2 Sample

The study required both a deep investigation and one that takes account of the rapidly changing assessment landscape in New Zealand. For this reason the investigation was restricted to three schools. This small sample, and the methodological approach itself means that the results will not be able to be generalised across all New Zealand schools, but in line with qualitative methodology, which emphasises *credibility* and *transferability*, the research will provide the thick descriptions needed for others to decide what transfers to their own schools and context.

It was important that at least one, but not all of the schools, had participated in the MoE's Assess to Learn (ATOL) professional development contracts. This is important as ATOL is the only substantial professional development programme in assessment offered to schools. Without including a school that had undertaken ATOL, it could be argued that any conclusions reached could be due to the absence of systematic professional development in assessment.

In order for the research itself to be manageable it was proposed that the selection of the schools be restricted to those in the greater Wellington region. However this did create problems. In the last 17 years, I have been principal of three Wellington

schools. I know many of the principals well and a number of them are close friends. In the interest of validity it is important that the schools in the study not include those where there is a close relationship between myself and the principal of the case study school.

1.5.3 Case Study Research

Chapter Five will also discuss the rationale for using a case study approach in more detail. In brief, it is important to make sure that the methodological approach matches the context of the research. Any research that seeks to illuminate what happens in schools needs to take account of the reality of school life. Schools are complex organisations with many competing demands placed on them. The implementation and management of assessment is just one of the many issues faced by school leadership teams that include curriculum implementation, pastoral care, behaviour management, personnel, financial and property management, and relationships with BOTs and parents. Schools attempt to meet these demands with a finite amount of resources and limited time frames.

Any project that seeks to understand how schools attempt to develop manageable and effective assessment systems begins with a choice between two main research paradigms that are loosely described as quantitative and qualitative. A small number of quantitative studies have attempted to shed light on the New Zealand assessment landscape. These included Dunn, Strafford, and Marston (2003) who investigated schools' use and rating of various assessment tools, and Wylie (1997, 1999) whose studies of the impact of the Tomorrow's School's reform included elements of assessment. However, the research questions in this study sought to develop a deep understanding of how assessment operates in the schools in question and how various key players attempted to make sense of the challenge of how to assess children in a manageable and effective manner. As such, qualitative research is the most appropriate approach in answering these questions. Mertens (1997) uses Denzin and Lincoln's (1994) definition to describe qualitative research:

Qualitative research is multimethod in focus, involving an interpretive, naturalistic approach to its subject matter. This means that qualitative researchers study things in their natural settings, attempting to make sense of,

or interpret phenomena in terms of the meanings people bring to them. (pp. 159-160)

Within the qualitative research paradigm, a case study was chosen as the most appropriate methodology to meet the needs of this research. In attempting to define the concept of a case study, Meriam (1998) stated “that a qualitative case study is an intensive, holistic description and analysis of a single instance, phenomenon or social unit” (p. 27). In this research, the phenomena under study are the manageability and utility of assessment, and the bounded systems are the three case study schools that are examined.

1.6 Structure of the Thesis

This first chapter has introduced the rationale for the research and context of the study. It has given a brief overview of the methodology. Chapter Two explores the concepts of manageability and utility. Attention is given to how these terms have been used in the literature. The relationship of manageability and utility to validity and reliability is explored. Chapter Two concludes with a definition of each of these two concepts.

Chapters Three and Four comprises the literature review of the thesis. Chapter Three traces government policy on assessment in New Zealand from 1989 to the 2006 and discusses these policies from the perspective of the major challenges to assessment that developed out of the implementation of the New Zealand education reforms and the impact that these changes have had on school practice. This chapter also discusses the major movements in western education in the last 25 years, covering outcomes based education, performance or authentic assessment, formative assessment, evidence-based teaching, and international comparisons of achievement. The discussion will include the development of the respective concepts, claims made in favour of each, and their impact in New Zealand. Chapter Four examines assessment tools and approaches developed in New Zealand since the reforms. They include NEMP, the ARBs, School Entry Assessment, STAR, The Exemplars, asTTle, and the revision of PAT Mathematics. While not a recent development, Running Records are a pervasive assessment tool and their use will also be discussed. The rationale for the development of each of these tools will also be examined and critiqued in terms of manageability and utility.

The research methodology will be discussed in more detail in Chapter Five. Chapter Six reports the results of the three case studies while Chapter Seven discusses how they relate to each other and the wider context of education. Chapter Eight concludes with an examination of the limitations of the research, and makes a number of policy recommendations for assessment practice in New Zealand as well as making recommendations for future research.

1.7 The Researcher's Perspective

This section outlines my own personal perspective on a number of issues that are central to the research. The research is interwoven with my own career as an educator which began in 1980 when I enrolled at the University of Canterbury to study education. During the following four years I completed a three year primary teaching programme at the Christchurch College of Education and a BA in Education at the University of Canterbury. This was followed by an M.Ed from the University of Canterbury which I completed in my first four years of teaching. At Canterbury I undertook a number of papers in educational measurement which mainly focused on norm referenced testing.

After teaching in state and integrated schools in Christchurch and the Wairarapa, I took up my first principal's appointment in 1993 and for the last 16 years have been a principal of three Catholic Schools in Wellington: St Anne's Newtown, 1993-1997; St Theresa's Karori, 1997-2003; and St Joseph's Upper Hutt, from 2003 to the present. My time as a principal has therefore largely coincided with the period being investigated in this research.

The fact that I am a practitioner-researcher, rather than an academic or policy developer has both advantages and disadvantages in undertaking this research. On the positive side, as a principal I have a comprehensive knowledge of assessment in primary schools. I report to parents, the Board of Trustees and the Ministry on children's achievement. I have experienced six ERO reviews. As a principal I see one of my main roles as leading teachers in developing assessment practices to enhance student learning. It is with irony, that I have to confess that while a motivation for my research has been to discover how schools can assess in a more manageable way, over the last five years this has led to the children at my school undertaking more

assessment than possibly happens in many other schools as I have sought to use, understand and critique all the assessment tools available to New Zealand schools by trialling each in some way or to some degree with various class groupings in our school. While these experiences, and my studies at the University of Canterbury in the 1980s, have given me a substantial background on assessment and how New Zealand primary schools operate, it not surprising that in over nearly 30 years involvement in education and 17 years as a principal, I have developed views, in fact strong views, on assessment that need to be acknowledged, indeed be suspended, in undertaking this research. From the very choice of the research topic it is plain that I have a particular perspective on assessment in New Zealand schools. I would not be researching the manageability of assessment if I did not believe that this was an issue worthy of investigation. By choosing to devote a substantial proportion of the last five years of my life to the topic I obviously believe it is an issue that needs addressing.

For the research to have validity it is important that my own positions are clear. In this research I am in one sense the *instrument of the research*. I identified the research topic, developed the research questions, gathered, analysed and interpreted the data. Where the researcher has such a dominant role it is important that they identify their own bias as they approach the study.

The following statements attempt to clarify the views I held on assessment in 2004 when I began this study. Since the introduction of the New Zealand curriculum I have believed that assessment against the levels of the New Zealand Curriculum is unwieldy, illogical and lacking research on children's learning to support such an approach. I believe that some form of organisation such as structuring the curriculum into levels is appropriate for planning purposes but does not provide a standard to assess children against. Developments that have attempted to assess students' achievements against the levels of the curriculum and the numerous achievement objectives have largely been a waste of teachers' time and produced little information that is useful.

Associated with this view was my belief that when compared with other elements associated with primary education, assessment is very technical and complex. Asking each school to develop their own system of assessment was not only impractical in terms of manageability it also asked individual schools to carry out a task that they

lacked the capacity to achieve successfully. It would have been more appropriate for the Ministry of Education to provide greater direction in the area of assessment.

A further belief that I held at the beginning of this research was the view that the development of children's ability in the core areas of numeracy and literacy was the most important activity primary schools engage in. In fact, primary schools are the only institutions solely charged with this responsibility. For this reason I held the view that assessment activity in primary schools should focus on these core areas. By devoting a large amount of teacher time to assessment in other curriculum areas, schools risk being distracted from their primary function in literacy and numeracy education.

I also hold the belief that while schools should be concerned with the development of high level attributes and values, such as found in the essential skills and the key competencies, these concepts are extremely difficult to assess. They are too complex and it is too difficult to separate the factors that are determined by home background and other influences from what are the result of the school curriculum. While it is appropriate for schools to set goals to develop children's' ability to *manage self* (MoE, 2007a) and to evaluate the effectiveness of their programmes in this competency, to extend this evaluation to include the formal assignment of scores or descriptors to individual students is, in my view, questionable use of the school's time.

While the views expressed above are concerned with elements of the New Zealand Curriculum Framework (NZCF) (MoE, 2003) I also began this research with two strong views that applied to assessment in more recent times. As already noted, I believed that the structure established in 1993 that asked schools to develop their own assessment tools (MoE, 1992, 1994b) placed serious stress on the school in terms of manageability. However, the fact that since 2000 there are now a large number of assessment tools available to schools has also put a great deal of stress on schools. In other words, schools in 1993 were placed under a great deal of pressure to assess learning within the curriculum with no tools to carry this out; thirteen years later the stress comes from having too many tools to select from in order to assess the learning.

In the last few years a number of professional development contracts have encouraged schools to devote what I consider to be an excessive amount of time on moderating the levels they assign to children achievement in writing. This is most common in the use of the asTTle writing tool. I believed that assigning a lot of time, to what in my view, was 'low stakes assessment' has not been a worthwhile use of teachers' time.

As stated, I have tried to suspend the views expressed above in analysing the data from the three case schools and let the participants and the schools speak for themselves.

As this research deals with the implementation of government policy, it is also important that in addition to my perspectives on assessment as a school leader I should acknowledge my own political perspectives. The period from 1989 to 2006 spans three government administrations: the Lange-Palmer-Moore Labour government of 1984-1990 that introduced Tomorrow's Schools; the Bolger-Shipley National government of 1990-1999 that introduced the curriculum reforms; and, from 1999 to 2008 the Clarke-led Labour government. I would describe my own political perspective as centre-left, and heavily influenced by a Catholic theology of social justice. These views have led me to be involved with the New Zealand Labour Party. As stated above, I believe it is important that in undertaking research that examines government policy, the political perspective of the researcher needs to be transparent. However, while acknowledging my political views I believe I have striven to be balanced in undertaking this research. In fact, this study was begun with the view that the assessment issues in New Zealand primary schools needed to be investigated from a pragmatic rather than an ideological perspective. This view was strongly stated in my research proposal. I have therefore been more interested in investigating how assessment can be made more manageable and effective in our schools rather than engaging in lengthy debate over ideological perspectives of the reforms and subsequent changes in education policy. However, the adoption of a pragmatic stance politically does not preclude the opportunity to conceptualise and theorise on the phenomenon being studied.

CHAPTER TWO: THE CONCEPTS OF MANAGEABILITY AND UTILITY IN PRIMARY SCHOOL ASSESSMENT

2.1 Introduction to Chapter Two

The aim of this chapter is to develop a comprehensive understanding of the concepts of the manageability and utility of assessment. It will begin by examining the use of the terms in the general assessment literature. The components of manageability will be examined and, where appropriate, examples of practice will be used to provide a clearer understanding of the concept. At their most fundamental level, manageability and utility concern the interaction between people, most often teachers, and assessment tools and systems. Considerable attention will be given to exploring the human dimension of manageability and utility.

While the major focus of this thesis is on the manageability and utility of assessment in the New Zealand primary school system, these terms will also be examined through the wider assessment literature and from sources related to other educational contexts. However, because these terms have greater meaning when they are considered in relation to the purpose for which the assessment is administered, this chapter will include discussion of the various purposes of assessment that are common within the context of primary school education. The chapter will then explore the concepts of the manageability and utility of assessment by discussing how they relate to the more established concepts of validity and reliability. The chapter will conclude with a definition of each term.

As a starting position, the term *manageability of assessment* will focus on concerns such as the time required to administer, score and record an assessment, as well as other practical matters such as the amount of time needed for training teachers to administer and interpret the assessment, and the costs associated with all of the above and the purchase of materials. A further element is the idea that a manageable assessment is one which does not distract too much from normal teaching activities.

Similarly, as a starting position, the term *utility* will be concerned with the effective use of assessment results. It matches what Baker, Linn and Herman (1996) term *the action element of utility*: “by action we mean the degree to which and how assessments are actually translated in practices and policies: how the assessment is actually used.” (p. 9). Or as they quote from a 10th grade history teacher: “How can I use this in my teaching?” (p. 9).

2.2 The Manageability of Assessment

2.2.1 Manageability in the General Assessment Literature

While the term *manageability* is relatively common in the literature and discussion of educational assessment, its research coverage appears not to be extensive. As is evident from the literature, manageability has important links with the concepts of validity and reliability. However, while the conceptualisation of validity and reliability is well developed in the assessment literature, based on long traditions of research, this is not the case with manageability. The term has been used more loosely and, as just noted, been subject to little research.

It is common, though not universal, for the concept of manageability, and related terms, to be covered in introductory undergraduate assessment texts. In these cases, the authors generally state that *manageability*, *usability*, *utility*, *efficiency*, and other associated terms are important considerations in designing and choosing assessment tools. However, the coverage is usually so brief it is questionable that the authors consider these concepts to be important. An example of this type of coverage of manageability in assessment is in the introductory text *Measurement in Assessment and Teaching* (Linn & Gronlund, 2000). The authors use the term *usability* to refer to matters that this chapter would describe as *manageability*.

In selecting assessment procedures, practical considerations cannot be neglected. Assessments are usually administered and interpreted by teachers with only a minimum of training in measurement. The time available for assessment is almost always in competition with other important activities for time in the school schedule (p. 133).

While it would be difficult to disagree with this view, the discussion of usability consists of less than two pages, while the book devotes 27 pages to validity and 25 pages to reliability.

Thorndike, in his introductory text *Measurement and Evaluation in Psychology and Education* (1997) uses the term *economy* to cover many of the issues other writers refer to as *manageability* or *utility*. While covering various aspects of assessment and evaluation, Thorndike's major focus is on psychometric testing. In discussing *economy* he includes the cost of test materials and the cost and ease of scoring. He also includes the time taken for administration of the test. However, this discussion is again relatively brief, covering seven pages compared to 34 on reliability and 50 on validity.

While Thorndike (1997) describes various *manageability* issues as *economy*, the term *practicality* is used by Ward and Murray-Ward (1999), and *practical considerations* by Sax (1997). There is something of a British Commonwealth versus American divide in terminology. While *manageability* is the most common term used by British writers (Connor, 1999; Whetton, 2004; Weedan, Winter, & Broadfoot, 2002) and in the New Zealand context (Allen, Crooks, Hearn, & Irwin, 1997; Elley et al. 2004; Hall, 2004; Hill, 2003; Knight, 2000), other writers, particularly American commentators, use the term *utility* (Baker, Linn, & Herman 1996; Messick, 1992).

2.2.2 Manageability and Utility

As is apparent from the above, it is clear that terms *manageability* and *utility* have been used in overlapping ways to refer to matters such as the time required for assessment, the cost of materials and the ease of administration. However, it is evident that at times the term *utility*, rather than *manageability*, has been used to refer to the use made of the results and their effectiveness for improving formative decision-making based on the information provided. In support of the distinction made between these concepts in the introduction to this chapter, the term *manageability* is used to include factors such as time, ease of administration, costs of materials, and the professional development needed to administer the assessment, whereas *utility* will be used to include how useful the information is to teachers, other personnel and agencies.

Two examples from current New Zealand assessment practice illustrate the distinction between the use of the terms manageability and utility in this writing. The 6 Year Net survey (Clay, 2002) is conducted close to a student's 6th birthday – while this practice is almost universal, it is not mandatory. The 6 Year Net consists of seven individual assessments: Running Records, concepts about print, letter identification, word identification, Burt Word test, 10 minute writing tests, and hearing and recording sounds. The whole survey takes on average over 2 hours per child to complete. As such, its manageability is open to debate. However, the results are used to identify which children are achieving well below their peers so that they can be placed on the Reading Recovery programme (Clay, 1993). In this way it could be argued the information has high utility.

The School Entry Assessment Kit (SEA) (MOE, 1997a) was introduced to New Zealand schools in 1997 and designed to be administered to children in their first six weeks of school. SEA consists of individual assessment in numeracy, oral language and emerging literacy. The estimated time for carrying out the assessment is approximately 45 minutes per child (Dixon & Williams, 2000). The comments of two writers on SEA highlight the differences between manageability and utility. Goldring (1999) investigated the use of the kit in Christchurch primary schools. In her interviews with teachers she identified major manageability issues. Comments about the time needed for administering the assessment appeared repeatedly. “Most teachers found it impractical and impossible to administer it in the classroom without the child and the rest of the class being compromised in some way” (p. 41).

Kelly (2000), writing from a ministerial policy perspective, is more focused on utility – what happens to the information?

Take for example School Entry Assessment, or any other diagnostic tool. Using the medical analogy of diagnostic testing, unless the evidence is analysed and the teacher is able to use the level of professional judgement required to put together the most appropriate programme to meet the student's needs; then conducting the test is not worthwhile (p. 70).

2.3 Components of the Manageability of Assessment

Within the search for a fuller understanding of manageability, it is valuable to examine how various authors have described the components that contribute toward

assessment manageability. Table 2.1 below provides the views of a selection of authors where manageability components are focused on formal, psychometric tests rather than assessment through classroom observation or assessment systems in general.

It is clear that all the four sources in the table emphasise the time taken to administer and score the test, the ease of administration, and the clarity of the test layout as important issues in the manageability of assessment.

Table 2.1: Components of manageability and utility emphasised in psychometric tests

Thorndike (1997)	Linn and Gronlund (1995)	Hopkins and Antes (1990)	Sax (1997)
1. Ease of test administration	1. Ease of administration	1. Frequency of test administration	1. Cost: both of test booklets and test scoring
2. A test has clear full instructions	2. Time required for administration	2. Time to allot for test administration	2. Time limits
3. There are few separately timed units and close timing is not crucial.	3. Ease of interpretation and application	3. Specific time for administration	3. Ease of administration
4. Clear layout of the test.	4. Availability of equivalent or comparable forms	4. Announcement of test date	4. Format
5. Interpretation of the scores	5. Cost of testing	5. Scoring procedures	5. Availability of alternate forms
6. A statement of the functions the test was designed to measure and of the general procedures by which it was developed.		6. Score interpretation	6. Multiple level examinations
7. Detailed instructions for administering the test.		7. Test assembly procedures	7. Scoring procedures
8. Scoring keys and specific instructions for scoring the test		8. Test directions	
9. Norms for appropriate reference groups		9. Test administration procedures	
10. Evidence of the reliability of the test			
11. Evidence on the intercorrelations of test scores.			
12. Evidence on the relationship of the test to other variables			
13. Guides for using the test and interpreting results within it.			

2.4 Manageability of Assessment Systems

The elements listed in Table 2.1 largely concern the manageability of individual assessment tools and, although involving a range of important considerations, are generally less complex than those relating to the manageability of an assessment system. An assessment system might involve national or state qualifications that use several assessment tools together. It could also refer to all the assessment policies, tools and strategies that are used within the one school. It is this use of the term *assessment system* that has most meaning for this thesis. However, the critique below illustrates the greater complexity usually found when dealing with the elements of an assessment system, in this case New Zealand's National Certificate in Educational Achievement (NCEA) administered by the New Zealand Qualifications Authority (NZQA).

In part it can be claimed that NZQA has been given an impossible task; the NCEA has all the hallmarks of being a bottomless pit for draining educational resources. The system is unmanageable: there are too many standards; too much assessment needed to support the standards; too many difficulties with setting examinations and tasks to meet vague predefined standards; too much time spent on cumbersome moderation procedures which do not work properly; too many difficulties in communicating clearly with teachers, students and other stakeholders; and too many manageability issues that require attention to detail beyond the capacity of NZQA to handle (Elley, Hall, & Marsh, 2004, p. 15).

While this statement concerns a secondary school qualification rather than an assessment system in a primary school, it does illustrate the point that when discussion is moved beyond evaluation of one particular tool to a whole assessment system with multiple tools and a number of different people involved in the process, the elements of manageability are more complex than with a single test. In contrast, examples of analyses involving the manageability and utility of individual assessment tools are seen in the evaluations provided within the *Mental Measurement Yearbook* (Young, 2004). This publication has a long history of critically examining individual tests and other assessment tools. Again an Atlantic divide is possibly evident in some of the reviews. For example, American writers tend to concentrate on individual assessment tools and the British, as a result of their experiences with Standardised Assessment Tasks (SATs) in the early 1990s, are more focused on assessment systems. As already noted, the term *assessment system* needs to distinguish between

the assessment that schools choose to undertake and assessment that is prescribed by a central agency. The following section looks at this distinction within the New Zealand primary school context.

2.5 Four Contexts relating to the Manageability and Utility of School Assessment

It is possible to identify four different contexts relating to school assessment: individual assessment tools, individual teacher assessment, school-based assessment, and prescribed assessment. The four contexts are to some extent arbitrary and overlapping. For example, prescribed and school-based assessment generally include a number of individual tools. However, a closer look at manageability issues in each of the contexts is helpful for clarifying some of the issues involved.

1. Individual assessment tools: This refers to tests or resources such as the Progressive Achievement Tests (PATs), asTTle, Concepts About Print (CAP), ARBs, and STAR. While it can be argued that the literature on the manageability and utility of assessment has been under-developed in the past, the context of individual assessment tools, especially formal tests, has had the most attention. The concepts of manageability and utility are frequently included in critiques of tests and draw upon the type of themes covered in Table 2.1. In relation to manageability and utility, decisions regarding individual assessment tools are generally based on choosing the most appropriate tool for the purpose of assessment. Decisions also relate to questions regarding the time required for administration and the time required for the training of teachers.

2. Individual teachers' assessment: This section refers to those assessments carried out in school classrooms on a day-to-day basis. Crooks (1988) included teacher-made tests, curriculum-embedded tests, oral questions, and a wide variety of other performance activities. This area might also include assessment of motivational and attitudinal variables in addition to the learning of skills. Individual teacher assessment also includes the internal “headnote” assessment that Hill (2000) describes and “log-in-the-head” as Timperley and Parr (2004) refer to it.

Individual teacher assessment is the most informal, but by no means least important, assessment carried out in classrooms. It includes such things as teacher observations, anecdotal recording, teacher-made tests, questioning and giving feedback. It is extremely hard to quantify the use of these assessment activities, yet it is crucial that they are taken account of in order to gain a comprehensive understanding of assessment. Manageability and utility issues concern trying to get a close alignment between this area of assessment and school-based systems.

3. School-based assessment systems: This refers to the assessment activity that is prescribed by the school management in its policies and procedures. A distinction is made here between the assessment activity prescribed by the central agency and that which the individual school decides to carry out. Here manageability and utility issues relate to questions such as:

1. How many and how often are assessment tools being used in the school?
2. Is there any unnecessary duplication between different assessment tools being used in the school?
3. How is the assessment information recorded?
4. What use is made of the information?
5. How much class time and teacher education is needed?
6. Does the assessment meet the purposes that were intended by the school and the teachers?
7. Do teachers have a clear understanding of the purpose of each assessment?
8. Is it time for the school to review its assessment system to ensure the manageability and utility of its assessment system?

4. Externally prescribed assessment: This section covers assessment or testing carried out in a school that is prescribed by a central agency or local governing authority. Prescribed assessment includes such things as national tests and individual testing that the school has no choice in carrying out. Where manageability and utility issues have serious implications they can only be addressed in changing the local, state or national policy through school and teacher advocacy groups and media campaigns. For example, Gipps (1994) and Whetton (2004) document how the British SATs were modified to become more manageable, if less authentic, through teacher

and media pressure. This will be discussed in more detail when examining the UK context of assessment.

2.5.1 The New Zealand Context of Prescribed and School-Based Assessment

While commentators have traced the reform of education in New Zealand in the early 1990s to the earlier British reforms, the New Zealand developments in assessment were quite different (Elley, 2004; O'Neil, Clark, & Openshaw, 2004; Willis 2002). The reforms did not include a national testing regime. Put most simply, and using the four contexts above, the prescribed assessment system in New Zealand was that schools were to develop and implement a school-based system. As already noted, the NZCF (1993a) did not provide the tools to assess learning against the new curriculum as required by the NAGs (MoE, 1993c). Schools were directed to develop their own assessment tools (MoE, 1992, 1993a, 1994b).

2.6 Manageability in Context: the UK Experience

In the last 20 years, one of the most notable examples of the importance of manageability in assessment was the controversial introduction in 1991 of SATs testing of 7 year olds in the United Kingdom. The requirements were that each teacher made an assessment of each pupil's level of attainment on levels 1-4 of the National Curriculum scale in relation to the attainment targets of the core national curriculum subjects. Teachers were to use a mixture of tests, practical tasks and observations (Torrance, 1995). The assessments took a minimum of forty hours per class and were rarely managed without support being available for the class teacher.

Gipps' (1994) description of assessment in science graphically illustrates what an absence of manageability looks like in actual classrooms:

Floating and sinking in science was assessed through a practical task in which the children were provided a range of objects and a large tank of water. The children had to predict which objects would float or sink and try to develop a hypothesis. Since it could take a week or more to assess a whole class of children on this particular task at one point in the summer term every infant class would seem to be full of water, waterlogged objects and rotting pieces of fruit. (p. 113)

This ambitious assessment system that sought to combine summative, diagnostic and formative assessment across a wide range of schools foundered on its lack of manageability. In 1991 the Scottish education authorities decided not to continue to participate in the assessment system after teacher protest. In 1992, assessment tasks were much reduced and increased use was made of paper and pencil tests. However, testing in that year took 30 hours of classroom time and in 1993 teachers actually boycotted the system. This was followed by an official review leading to the *Interim Report on the National Curriculum* (Dearing, 1993) which described the assessment system as “unwieldy”. In response to the Dearing Report the government restricted assessment for 7 year olds to literacy and numeracy and increased the use of paper and pencil tests at the expense of more practical and authentic forms of assessment. The movement from an ambitious and comprehensive but ultimately unmanageable system, which combined summative, diagnostic and formative assessment with large elements of what was described as authentic assessment, has been well documented (Broadfoot, 1995; Elley, 1993, 2004; Gipps, 1995; McCallum, Gipps, McMaster, & Brown, 1995; Torrance, 1995).

2.7 Manageability is about People

A defining feature of manageability and utility is that they concern people. Baker, Linn, and Herman (1996) used the term utility when they wrote insightfully about the human dimension in assessment:

The heart of our model of utility is the person, the human dimension, not abstract methodology, a particular analytic technique or any preferred format or test. For it is, after all, a person who must make inferences that are accurate, fair and appropriate for the particular purposes and students. People make judgments about whether they can trust, can understand, and will use information. People take action or do nothing; their choices of action fit to real limits of available knowledge, sense of benefit and understanding of costs. Creating better methods, high precision techniques, more inclusive assessments, and glossier, high tech reports of results is of little use unless people use assessment results wisely to achieve worthy goal. (p. 10)

This concept was illuminated by Pole (1993) in his study of the implementation of records of achievement in one British secondary school. Records of achievement were introduced into a substantial number of British schools in the 1980s. Their use was recommended, rather than mandatory and many schools discontinued with the system when a very demanding assessment regime was introduced following the 1988 Education Reform Act. Records of achievement involved negotiated goal setting

between teachers and students. It involved, among other things, detailed one-to-one learning conversations between teachers and students termed negotiation sessions.

The recording process should help schools to identify the all-round potential of their pupils and to consider how well their curriculum, teaching and organisation enable pupils to develop the general practical and social skills which are to be recorded. (p. 2)

In his research, Pole (1993) uncovers the veil of the working lives of teachers and school managers. Quoting the deputy head of the school:

I mean to be honest... only half the staff were for it. Now to persuade some of them, the main bugbear they saw was time. Now we convinced them, I believe, because we said to start with you won't have to do reports because records of achievement were going to replace reports. (p. 20)

The implementation of the records of achievement at the school in Pole's study was a constant challenge in terms of the time required to administer them. The major negotiation sessions with pupils took on average 30 minutes each. While most of the teachers believed that in terms of utility the innovation had many positive benefits for teaching and learning, they felt the system had significantly impacted on their workload and ability to perform other functions. After five years the approach had become part of the bureaucratic functions in the school. That is, there were deadlines for completion of aspects of the task and a major goal of teachers became to meet the deadlines, complete the negotiation sessions in 30 minutes, and not hold up other teachers' use of the negotiation room. The process became focused on the completion of the target setting and negotiations as something to be ticked off, while the ongoing revisiting of the goals in learning conversations between the teacher and the pupil during the year, which was in many ways the heart of the process, was neglected.

Pole's (1993) findings were echoed in a recent New Zealand study by Lovett and Sinclair (2005) of the assessment activities carried out by beginning teachers. Lovett and Sinclair's description of Danny shows a teacher whose assessment focus is concerned with complying with the school schedule rather than making the process integral to his teaching.

It seemed for Danny especially, the daily grind of teaching and perhaps his unpreparedness for assessing children's learning was causing him to follow the school's schedule rather than develop his own momentum and decisions for classroom assessments. He followed the school's assessment requirements and this took all his energy. Over the five months of this study, Danny's talk about assessment differed from his practices. It was becoming increasingly

apparent that his assessments were tacked on to the edge of his teaching for compliance purposes rather than to direct his teaching. He was fully stretched to keep pace with the assessment in his school. (p. 51)

Both Pole's (1993) UK study and Lovett and Sinclair's (2005) New Zealand study capture what is problematic for many elements of school assessment. The more basic elements of the assessment process such as carrying out the assessment and recording the results in the timeframe required by the school are compliance requirements. However, what many commentators see as the most important element of assessment, such as the use of assessment data to inform teaching, are largely voluntary activities by the teacher; they largely happen behind closed doors and rarely leave a paper trail.

Whether it is voluntary or not, the problem with attempting to research a great deal of the real assessment teachers do is that it is not visible. In discussing the problem with what she termed headnote assessment, Hill (2000) stated:

Due to 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. (p. 253)

A number of the teachers in Hill's study were operating two systems of assessment. They were meeting the school's requirements in terms of carrying out prescribed assessments and record keeping. However, these requirements did not give the teachers what they considered was a valid and useful understanding of their students' learning. That is, for the teachers themselves the school assessment requirements had low utility. For this reason they were also carrying out their own internal "headnote" assessment. As Hill stated, "many teachers expressed frustration at recording assessment in line with school policy while, at the same time, operating their teaching according to the learning progressions they have in their minds." (p. 10).

In a similar fashion, Symes and Timperley (2003) report the challenges of a South Auckland school trying to move from merely recording assessment information to making more analytic use of assessment data. At the beginning of the study teachers were very candid on their lack of attention to formal assessment information. While all teachers identified test results as being part of the information systematically collated for school use, they saw this as solely the interest and responsibility of the assistant principal, who administered the procedures. The information provided had, in the view of the teachers at the start of the study, little utility in terms of their

teaching. On receiving the information they admitted giving the results little more than a cursory glance. One even admitted to having never actually looked at the results sheet because it had been filed away by a teacher aide. Teachers saw little reason in their busy schedule to refer back to these records and admitted once they were filed away, they were rarely looked at again. “They go into the file but you don’t have time to look at them again”. (p. 37)

Whetton (2004), writing in a British context, provides a different perspective on manageability. He challenges the claims by teachers and the media that the SAT assessment system is unmanageable:

From its very inception, there had been opposition to National Curriculum testing, but this now started to become even greater. The phrases “The Most Tested Pupils in Europe” or “The Most Tested Pupils in the World” began to appear in articles advocating reductions in the amount of testing. Teachers and educational commentators agreed with this and consistently used such arguments. The tests are constantly referred to as “flawed” and pressure groups such as “Authors against the SATs” sprang up, receiving media attention. (p. 11)

Yet Whetton (2004) points out that over the period of nine years of education, English children only spend 17 hours undertaking compulsory testing.

In a similar vein in their evaluation of teachers’ use of the New Zealand ARBs, Gilmore and Hattie (2001) found that the resource was only being frequently used by three percent of New Zealand teachers. This research illustrates the relationship between utility and manageability where teachers have different perceptions of an assessment tool. While time was considered a barrier by low and non-users, it was not an issue for teachers who made frequent use of the tool.

Another perceived barrier was time - time to access the computer, search ARBs, and download and print resources. The workload of teachers is far from trivial. This suggests that for ARBs to be used or accessed more often the perceived importance of the resources as excellent assessment tools for classroom practices needs to be addressed. It is noted, however, that “time”, or lack of time, was more of a problem for the non-user than for users. Clearly, the users saw the returns for their investment in time, whereas non-users cited time as a major reason for not using ARBs. It is priorities for time rather than time that is a critical factor. (p. 254)

A further study that illustrated the complex nature of teachers' perceptions about assessment was carried out by Brown (2002). In this New Zealand study, Brown identified four major and relatively stable perceptions individual teachers held about the purposes for assessment.

1. The purpose of assessment is to improve teaching and learning.
2. The purpose of assessment was to identify if children had made progress, achieved in the learning. Brown uses the term: *certification of learning* to describe this concept.
3. The purpose of assessment is for the accountability of schools and teachers.
4. The rejection of assessment; assessment is irrelevant to the life and work of teachers and students.

Brown (2002) concluded that teachers see assessment as irrelevant when they have been in situations where a large amount of assessment data has been collected for accountability purposes and little use has been made of the information. This view is also supported by Symes and Brown (2003).

2.8 The Purpose of Primary School Assessment

As stated at the beginning of the chapter, the concepts of the manageability and utility of assessment take greater meaning when they are discussed in relation to the purpose of assessment. Within the context of school assessment, various authors have documented a range of possible purposes covering different assessment tools. While much of this writing shares a number of common themes such as informing teaching, reporting to parents and matching students to appropriate materials and programmes, there has been a shift in emphasis over time, probably reflecting different philosophies of education and the focus of particular assessment tools. For this discussion, a preference has been given in this thesis to New Zealand sources in order to reflect the context in which this research is based. The discussion will also focus on assessment at primary level rather than on secondary and tertiary education where certification of learning (MoE, 1994b) is a major purpose for assessment.

The Progressive Achievement Test reading tests (PATs) introduced to New Zealand forty years ago (Elley & Reid, 1969) and revised in 1991 (Reid & Elley, 1991), list eight purposes for this standardised tests that were representative of that period. The

purposes of the readings tests were recorded by Reid and Elley in their 1991 revision as:

1. To group children with similar needs and competencies within a class
2. To select realistic goals and select appropriate materials for all students
3. To identify readers requiring special diagnostic and remedial treatment
4. To select able readers to be considered for special enrichment programmes
5. To identify students who are not working up to expectation, or who are performing erratically
6. To provide a frame of reference within which teachers can evaluate the class work of their students
7. To locate areas of weakness and strength within a class
8. To help classify school entrants from outside the New Zealand school system (pp. 25-27).

The recent revision of the test (Darr, McDowall, Ferral, Twist & Watson, 2008) includes most of the eight purposes above as well as including “measuring progress over time”; with the inclusion of a reading scale which was not a feature of earlier PAT tests. It should be noted that points 6 and 7 above are expanded to include examination of performance on different texts and different test items. These elements appear to reflect the movement towards greater analysis of assessment data associated with the recent trend to evidence-based teaching. This movement will be discussed in Chapter Three. The 2008 edition of the PAT Reading tests also includes reporting of results to parents and caregivers, reflecting the increased pressure for accountability to parents over the last two decades. It is also interesting to note that purpose eight, “the classification of school entrants from outside the New Zealand school system” is no longer included, suggesting a more sensitive approach to assessing students for whom English is a second language.

Writing more generally about the purposes of assessment at the start of the educational reforms the MoE (1994b) lists four main purposes of assessment in its publication *Assessment policy to practice*:

1. Improving learning

To identify the knowledge and experience which students bring to a learning task, to plan and /or refine teaching and learning programmes and to meet individual and group needs.

To monitor students' progress and to gauge what learning is actually taking place. Teachers can use assessment tools to find out whether expected learning is occurring (for example, whether the specific objectives set by a teacher within the broad achievement objectives in the national curriculum statements are being met), and whether any difficulties are being encountered.

Assessment can be used to provide feedback to students.

2. Reporting progress

Information gained through assessment of students' progress and levels of achievement can be provided for parents.

Assessment can be used to provide transition information when a student moves to the next stage of learning.

3. Providing summative information

Assessment can be used to build a profile of a student's achievement across the curriculum – her or his knowledge, skills and attitudes.

Assessment can help students make decisions about future study options

4. Improving programmes

Assessment can provide information for the evaluation and review of programmes throughout the school. It can help teachers determine which programmes are effective, how students are progressing and where improvements may be needed.

(pp. 7-8)

What is surprising about the purposes stated in *Assessment policy to practice* (MoE, 1994b) is the lack of attention given to assessment for accountability purposes that was a strong element in the outcomes based education (OBE) structure of the reforms of the early nineties. This point will be discussed further in a Chapter Three. However, through the use of expressions such as *learning task*, and *the specific objectives*, the document reflects a view of education that appears to be more concerned with discrete units (achievement objectives) within a highly structured

curriculum rather than viewing learning as a continuous process; with the focus on “the learner” rather than the curriculum common in New Zealand education prior to the reforms. This element will also be discussed in more detail in subsequent chapters.

In their guide to parents, ERO (1996) includes reporting to BOTs as a purpose of assessment so that appropriate interpretations can be made. They also include the setting of achievement targets as an assessment purpose. While the term *achievement targets* is similar to *realistic goals* used prior to 1989 (Reid & Elley, 1991) it reflects a shift to the managerial language associated with the reforms.

In their argument for the introduction of national testing, ERO (1998) stated clearly that a major purpose of assessment, in their view, was the accountability of schools to government and to parents. More recent statements by the MoE (2004) also acknowledge accountability as a purpose of assessment:

- Provide assurance about the quality of education
- Provide the means of evaluating the progress towards raising achievement and reducing disparity.

In summary, the main purposes of assessment in the context of New Zealand primary school education appear to be: informing learning; reporting to parents; reporting to BOTs and the government; the evaluation of programmes; and informing the strategic direction of the school.

2.9 The Relationship between Purpose, Validity, Manageability and the Utility of Assessment

The relationship between the manageability, utility, and validity of assessment is complex. The following discussion attempts to illustrate the concepts and their relationship in relatively simple terms. In attempting this illustration an essential question to explore is whether the manageability of assessment is a separate and fundamentally different concept from validity or merely an element of validity in the same way that concepts such as content validity, construct validity and criterion-related validity are different aspects of the concept. The same question also needs to be explored with the utility of assessment.

Stated simply “a valid assessment task is one that fulfils its intended purpose” (Hall, 2008, p.6). Miller, Linn and Gronlund (2009) explain validity in terms of four major considerations:

Content Validity: How the assessment tasks represents the domain of tasks to be measured.

Construct Validity: How well the performance or assessment can be interpreted as a meaningful measure of some characteristic or quality.

Assessment-criterion relationships: How well performance on the assessment predicts future performance or estimates current performance on some valued measures other than the test itself.

Consequential validity: How well use of assessment results accomplishes intended purposes and avoids unintended consequences. (p. 74)

It is the concept of consequential validity that has the closest relationship to manageability and utility. Miller et al (2009) cite Messick when they state:

Messick (1989, 1994) has argued persuasively that an overall judgement regarding the validity of particular uses and interpretations of assessment results requires an evaluation of the consequences of those uses and interpretations. Assessments are intended to contribute to improved student learning. The question is, do they? And, if so to what extent? What impact do assessments have on teaching? What are the possible negative, unintended consequences of a use of assessment results? (p. 95)

In discussing consequential validity, Hall (2008) notes the following examples of consequential validity:

- The extent to which assessment tasks avoid harmful side-effects, such as creating unnecessary stress in students or promoting surface learning instead of deep learning (examples of *consequential validity*)
- The extent to which the marking procedures and feedback processes help students improve their future performance (an aspect of *consequential validity*). (pp. 6-7)

If a valid assessment task is one that “fulfils its intended purpose” and, as noted by Hall (2008), consequential validity is the extent to which an assessment procedure avoids unintended negative consequences, there is strong implication that the manageability elements discussed at the beginning of this chapter (see Table 2.1) are factors that lie within the notion of consequential validity. For example, Table 2.1 included ease of test administration, time required for administration, and ease of interpretation and application. All of these, if not carefully addressed in the design of

a test, will have negative backwash effects on teaching and/or learning. As noted in section 2.7, the literature suggests that unless assessment is manageable, teachers do not feel that they have the time to look deeply at what the assessment data provides.

However, as also noted in this chapter, there are important management features that relate to the wider context of assessment that are not directly related to validity of an assessment tool. These elements include:

- How much other assessment is being carried out in the school
- The other demands being placed on teachers
- Individual teacher's perception of the particular assessment tool. As discussed earlier when examining teachers' use of the ARBs, Gilmore and Hattie (2002) noted that low users of ARBs perceived the assessment procedures as being too time-consuming whereas high users of the resource believed the resource to be very worthwhile and did not consider the time required to undertake the assessment a problem. The importance of teachers' perception is also illustrated in teachers' views on the utility and manageability of the school's assessment procedures (Hill, 1999), or assessment in general (Brown, 2001).

The relationship between what has been described in this chapter as the utility of assessment and consequential validity is even closer than in case of the manageability of assessment. Miller et al's (2009) definition of consequential validity as "how well use of assessment results accomplishes intended purposes and avoids unintended consequences" (p. 74) largely covers the same intention as has been used to describe the utility of assessment in this Chapter. However, as with the manageability of assessment, there are elements of the utility of assessment that are external to the original purpose of the test. To illustrate this point, assessment procedures such as PATs and Running Records have high utility when used for their intended purposes. However, as will be described later in the thesis, these assessment tools were often used for purposes for which they were never designed. For example, principals in the early and mid-nineties would use them to report student achievement to the school's BOT in terms of the new curriculum and ERO expectations; however such data were not sufficiently well aligned with the particular achievement objectives and levels of the curriculum that were being assessed. .

In summary manageability and utility have a direct relationship with validity, in particular, consequential validity. If an assessment is to meet its purpose and avoid unintended consequences it has to be reasonable in terms of demands regarding ease and time devoted to administration and marking. In this sense the manageability and utility elements of the individual assessment tool or procedures should be considered to be sub-elements of the larger concept of validity. However, when assessments are considered in the wider context of both the individual school's assessment system or any external requirements of local or national bodies, other elements of manageability need to be considered. These include what other assessment activities are being carried out, other demands being placed on teachers, and teachers' past experiences and perceptions of assessment. While these important elements of manageability may impact on the validity of an assessment, they need to be considered as external to the concept of validity.

In a similar manner, the concept of the utility of assessment has both direct and indirect links with the concept of consequential validity in that the results of testing and the data obtained should relate to the purpose for which they were collected and provide information that is useful, not harmful, to students and teachers. The indirect or external links emphasize that when assessment data are used for different purposes or are embedded in a context that extends the intended application of the test or assessment the utility of that information may not be high.. In the first situation, utility aligns with "fitness for purpose" – if the results are useful as intended, they are fit for purpose. In the second situation, the assessment data may be used successfully for different purposes, but clearly a good deal of caution is needed to ensure that the data are fit for the new purposes and do not result in unintended consequences that would not normally be an effect associated with the test. Such a situation might arise, for example, when data collected for formative purposes (to assist decision-making about each student's next learning step) is aggregated and used for school accountability. A low decile school, for example, might well face hard "accountability" questions yet be successful in the formative sense.

2.10 Definitions of Manageability and Utility

It is clear that the manageability and utility of assessment is a significant issue for schools. While many writers have illustrated the negative effects of manageability in relation to assessment systems, few actual research studies have been undertaken.

Understanding of assessment would be enhanced if writers in the field used a common term to describe collectively such issues as the time taken to carry out assessment, marking, the amount of training required to administer the test and interpret the results, the cost of testing booklets, and ease of scoring. A number of different terms are presently being used: manageability, utility, usability, efficiency and economy. Of the various terms manageability is arguably the most appropriate because it is the most widely used and conforms to the common use of the word. It is not always easy to separate the manageability of assessment from how well assessment information is used; however, it would be helpful to make this distinction.

The concept of the manageability of assessment is dependent on the context in which the assessment is carried out and the perceived manageability of the activity in the eyes of the person or people who are describing the assessment. For this reason the concept of manageability involves a judgement that is both subjective and context bound. However, the subjective nature of the concept of manageability does not detract from its importance in any discussion on the development and implementation of an assessment tool or system. For the purposes of this thesis, which focuses on primary school assessment, manageability is defined as:

The judgment made about the ease or difficulty of carrying out an assessment activity. The judgement may focus on such factors as the time taken to carry out assessment, the time required for marking and analysis of the results, the amount of training required by personnel for administration, and the complexity of the assessment materials. The manageability of an assessment activity is dependent on the context in which the assessment is carried out and the perspective of those people carrying out the assessment.

This study examines both individual assessment tools and the wider concept of assessment systems as they relate to individual schools and education systems. It is therefore helpful to expand the above definition of manageability by describing the characteristics of a manageable assessment system. A manageable assessment system is one which allows a school to carry out its assessment and reporting requirements effectively, while at the same time is balanced in terms of the demands it places on all school staff and the time allocated to teaching. The manageability of assessment

concerns the interaction between assessment tools (tests, observations, and performances) and the people who use those assessment instruments. As such the manageability of assessment involves subjective judgements about the meaning and value teachers and school leaders make of the assessment activities they carry out.

In other words, an individual assessment tool is manageable where it effectively meets the school requirements, without placing excessive demands on the school in terms of the time required for administration and the processing of information as well as the training of staff.

The concept of manageability of assessment is clearly related to the utility of assessment. By utility it is meant that the assessment provides useful information; that is clearly a major element of “fitness for purpose”, a defining characteristic of a “valid assessment.” Utility can be interpreted for both individual assessment tools and for school-wide assessment systems. Quality of information and its use are key elements in both cases. For school-wide assessment systems, the major focus is not only on how much and how time-consuming is the gathering of the information, but the use of the assessment information to the school. For the purposes of this thesis the utility of assessment is defined as:

Judgements made about the quality and usefulness of the information gained from the assessment activities. Any consideration or research on the manageability of assessment needs to make clear the context in which the term is being used. Is the discussion focused on the manageability of an individual assessment tool or the manageability of a school, district or national system? The analysis of the manageability of an individual assessment tool is very important but relatively straightforward. The manageability of assessment systems is considerably more complex but deserving of a great deal more attention than it has had in the past.

As stated above one of the most challenging aspect of examining the manageability and utility of an assessment system is that it concerns the interaction of people assessment procedures. As such it concerns subjectivity and the meaning people make of the activities that make up their professional lives, and is therefore, extremely complex. However, the fact that this is challenging and complex does not mean that we should avoid progressing with attempts to further understand these issues.

2.11 Summary Chapter Two

Figure 2.1 is also an appropriate tool to summarise this chapter. As well as seeking to explain the relationship between manageability, utility and validity, the diagram illustrates the rationale for this research. It is argued that the manageability and utility of individual assessment procedures has had little attention in research and other writing on assessment. However, it is the elements in the outer circle of the diagram that are a particular focus for this research. That is, an exploration of the relationship between assessment requirements and practices with issues of manageability and utility within context of the New Zealand primary education system during the period 1990 – 2006.

CHAPTER THREE: ASSESSMENT IN NEW ZEALAND PRIMARY SCHOOLS 1990-2006

3.1 Introduction to Chapter Three

Chapter Three provides the context and background to answer research question one later in the thesis:

What were the major influences and challenges to assessment in New Zealand primary schools since 1990?

The chapter begins with an outline of assessment practices in New Zealand schools prior to 1989. This will be followed by an examination of Outcomes Based Education (OBE) which was a major influence on the reform of New Zealand education. The discussion on OBE will be followed by Table 3.1, which provides a chronological overview of issues, initiatives and publications related to assessment in New Zealand primary schools during the period 1989 to 2006. This chronological overview will be followed by a brief discussion of the major themes and challenges arising from the reforms in terms of assessment practices in schools. This discussion will identify four major challenges to assessment in New Zealand primary schools that resulted from the structure and processes initiated by the reforms:

1. The lack of direction given to schools in terms of assessment practices and the initial lack of the provision of assessment tools for the new curriculum.
2. The structure of the curriculum in terms of eight levels and numerous achievement objectives.
3. The lack of distinction in the NAGs between the core curriculum of literacy and numeracy and the other areas of learning.
4. The impact of the Education Review Office.

The response of schools to each of these challenges will be discussed as well as looking at the more general response of schools to the reforms. Changes in government policy since the millennium will also be discussed.

While the focus of this chapter is on New Zealand developments in assessment, to examine this period of the country's education history from solely a national perspective would be misleading. While the accountability movement in many ways led to the initial reforms, during the last twenty years four other major assessment movements, common to most western countries, have also influenced assessment practice in New Zealand primary schools during the period being examined:

1. Authentic or performance assessment
2. Formative assessment
3. Evidence-based teaching
4. International and national testing

These forces will be examined in terms of their influence on assessment practice in New Zealand and challenges they presented in terms of manageability and utility.

3.2 Assessment Activity in New Zealand Primary Schools in the 1980s

Prior to the reforms, assessment in New Zealand primary schools was low-level and non-obtrusive. In fact, texts that examine primary education in New Zealand in the twentieth century give little attention to assessment after the abolition of the proficiency examination in 1936 (Codd, Harker & Nash, 1990; Lee, 1980; Marshall, Coxton, Jenkins & Jones, 2000; Mitchell, 1968). The most notable assessment requirement in the decade prior to the reforms was the Progress and Achievement (P and A) Registers. All teachers were required to maintain P and A Registers that rated children's achievement on a five point scale on aptitude, attitude, and achievement in a number of curriculum areas. These records were not shared with parents, the children themselves or any outside agency. Their main use was to provide the child's next year's teacher, or in the case of someone transferring to another school, their new teacher some background on the child. P and A registers were compulsory, but were also low-stakes assessment and required little time commitment. The five point scale was in the process of being replaced at the end of the eighties by a criterion-

referenced system that was based on written descriptions of what children could do, rather than the norm-referenced five point scale of the P & A system (Hughes, 1982; Smith, 1982). However, these changes were overtaken by the education reforms (Crooks, 2002).

In addition to P and A Registers, a large number of schools annually administered the NZCER-constructed PAT tests in Reading (Reid & Elley, 1991), Mathematics (Reid, 1993), Listening (Reid, Johnson, & Elley, 1994) and Study Skills (Reid, 1977). While all of the tests are norm-referenced, the PAT Reading Tests did include some criterion-referenced information. In contrast to P and A registers, PAT testing was optional. The results were not analysed in any detail and it was very rare for the scores to be shared with parents. PAT results were used to group students for instruction and identify the gifted and those who would require more support. By the 1980s, a number of schools, especially those with a high percentage of Maori and Pasifika students, had ceased to carry out PAT testing on the grounds that they were biased against minorities (Consumer, 2002; ERO, 1999; Reid & Gilmore, 1983). The debate on cultural bias and standardised testing focused on both PAT tests and NZCER's now discontinued Test of Scholastic Abilities (TOSCA) and received full coverage in the *New Zealand Journal of Education Studies* in the early eighties (Beck & St.George, 1983; Nash, 1983; Reid & Gilmore, 1983; Shuker, 1987; Tuck, 1983). The overall impact of the debate was to lessen schools' reliance on PAT tests for obtaining external assessment information.

The most comprehensive form of assessment carried out in schools was the battery of tests known as the *six-year nets* (Clay, 2002). These individually administered tests were taken as close as practicable to a child's sixth birthday and were carried out to select those children who were to go on the Reading Recovery programme (Clay, 1985). By the late 1980s, Reading Recovery was, if not universal, very widespread in primary schools. In order for schools to be funded for Reading Recovery, the six-year net had to be undertaken. The taking of Running Records was a significant innovation in primary school assessment during the 1980s. Running Records were based on Clay's (1985) work in Reading Recovery. At middle and senior levels of the school, a very similar practice was referred to as *Informal Prose Inventories*; the development of those tools was based on the work of the Goodmans (1972) and the whole language approach to the teaching of reading and writing. Both Running Records and informal

prose inventories involve a teacher listening to a child read a series of graded texts; the child's reading is then scored based on the percentages of words read correctly. These percentages are translated into reading ages according to the grade (reading age) of the text.

The prevailing approach to the teaching of writing was based on the *process writing* method (Graves, 1983) that swept the country during the 1980s. Fundamental to process writing was the *writing conference* where the teacher and the pupil held a discussion based on a selection of the child's writing so that further developments could be made. The competencies the child had developed were noted and the acquiring or practising of the next most appropriate skills was discussed. In this way, the writing conference was an explicit example of formative assessment twenty years before the term became dominant in education.

In mathematics, assessment included school-developed basic facts tests that prevail in many schools today. Pre-tests were also common in mathematics. In junior classes, the Beginning School Mathematics resource (BSM), which had been introduced in 1986, included a large element of performance assessment (Visser & Bennie, 1996).

Formal and informal tests of spelling, such as Daniels and Diack (1958), Peters (1979) and Schonnell were common (Croft, Strafford & Mappa, 2000), as were handwriting samples. Reporting to parents largely consisted of one or two written reports and parent interviews each year. There were some elements of self- and peer-assessment in individual classrooms.

It may be helpful to summarise the differences in assessment activity between the decade of the 1980s and the last 15 years by stating what was missing from the earlier period that are common practices now. Assessment in the non-core curriculum areas was either non-existent or at a low level; the main goal was to have some information for reporting to parents and completing the P and A registers. Little or no analysis of data from any assessment was made. Portfolios may have been developing in a very small number of schools. Reporting to parents formally through aggregated data, such as happens now with BOTs, did not occur. In summary, assessment was low key and unobtrusive.

James (2006) stated that to fully understand assessment activity, there is a need to trace the links between particular learning theories, learning activities and assessment.

James also acknowledges that it is common and even appropriate for teachers to be operating multiple theories of learning. If assessment in the decade preceding the implementation of the reforms is examined, a number of learning theories can be observed. For example, basic facts tests in mathematics, letter identification in the six-year nets, and much of the assessments of spelling were based on a behaviourist view of learning. James describes a behaviourist approach to teaching as one where the teacher's role is to teach students to respond to instruction correctly and rapidly. On the other hand, the performance assessments of BSM were based on a constructivist view of learning. The student's prior knowledge was important and the child's ability to manipulate material in a certain way demonstrated their internal understanding and the schema they brought to the task. On a different tack, the conferences associated with process writing were examples of assessment based on what James terms a socio-cultural theory of learning. Although Graves (1983) does not link the process writing approach to Vygotsian theory, the writing conference had many of the characteristics of such an approach. There was an apprentice–master relationship and teachers were encouraged to model their own writing. Learning was socially constructed and involved peer interaction and criticism.

In summary, prior to the reforms, assessment was low-key and drew upon, inferentially or explicitly, a range of different learning theories or beliefs.

3.3 The Accountability Movement and Outcomes-Based Education

OBE has been used to describe the raft of changes associated with the education reforms in New Zealand 1989-1993. From the 1970s, movements began throughout the western world that challenged the pervading Keynesian economics and the structure of the welfare state. A number of terms have been used to describe these movements. They include *Neo-Liberalism* (Weiner, 2007), *The New Right* (Peters & Marshall, 1996; Sullivan, 1992) and *Monetarism* (Gordon, 1991). All these terms cover similar views of government services that had a dramatic effect on education policy and practice.

Simon (1993), writing about the causal factors in Britain for this dramatic change in education practice, traces one of the reasons for the rise of a monetarist approach in education to the economic events of the seventies.

Certainly the industrial, financial and state fiscal crisis of 1972/1973 marked a crucial phase in the introduction of new policies in education (and elsewhere). The fiscal crisis provided the material ground for monetarist policies which of course, particularly afflicted the social services whose expenditure must be tax raised. Keith Joseph, perhaps the leading Tory monetarist in the late seventies, consistently stressed the extent to which such expenditure reduced resources available to the wealth producing sectors of the economy. Educational expenditure, in this scenario, was increasingly seen as an unwelcome drain on productive investment and activity – as parasitic. (p. 30)

Tomlinson (2007) also writing from a British perspective, saw the impact of monetarist or neo-liberal policies in education as the breaking down of a consensus in education policy that lasted from 1945 to 1979 that was based on the idea “that the governments should regulate and resource education to achieve some measure of social justice and provide equal opportunities for all” (p. 173).

The influence of neo-liberal policies in education has also been acknowledged in other descriptions of developments over the past 20 years. These include: *the accountability movement*, (Sleeter, 2007); *outcomes based education* (OBE) (Hall, 2007; Tomlinson, 2007); and *managerialism* (Alphonse, 1999; Knight 2000). In this section, OBE will be the dominant term used because it has a strong focus on assessment.

3.3.1 The Development of the Accountability Movement and OBE

As stated above, from the 1970s onwards there was a growing demand in western economies for increased accountability in all public sectors including education. This can be traced through the OECD (1983) both in a New Zealand context and internationally (1995). Dixon (1999) cites Ewing (1977) who identified increased frequency in the use of the term *accountability* in education discussions during this decade. Dixon also traces the loss of public confidence in the capacity and will of education professionals to undertake effective education reform. This movement led to a reconstruction of education in general, with particular emphasis on how schools operated, their governorship, curriculum and assessment.

Writing in an American context, Ward and Murray-Ward (1999) trace demands for accountability to the same time-frame.

The 1970s saw a shift in focus from the federal to the state level and the advance of the accountability movement and of minimal competency testing.

During this period, an increasingly disgruntled public, impatient with what were perceived as small educational gains, demanded clear “evidence” of educational attainments. (p. 52)

The culmination of the accountability movement in American education was the *No Child Left Behind* (NCLB) Legislation in 2001. NCLB is federal legislation that designates specific outcomes for schools systems and states in America. Sleeter (2007) summarises the purpose of the legislation as follows:

The NCLB mandates that states receiving federal funding: To implement statewide accountability systems covering all public schools and students. These systems must be based on challenging State standards in reading and mathematics, annual testing for students in grades 3-8, and annual statewide progress objectives ensuring that all groups of students reach proficiency within 12 years. Assessment results and state process objectives must be broken out by poverty, race, ethnicity, disability and limited English proficiency to ensure no group is left behind. (p. 20)

The first large-scale legislative evidence of OBE policy being put into practice was the 1988 UK Education Reform Act. A number of authors (Elley 1993, 2004; Gordon & Whitney 1997; Sullivan, 2002; Thrupp, 2005; Willis, 1992) consider the New Zealand reforms in education to be derived from the UK 1988 Education Reform Act.

3.3.2. The Nature of Outcomes Based Education

OBE has been seen as the process by which the accountability movement philosophy of education has been put into operation. This approach to education is concerned with more than just the teaching of the curriculum and assessment of learning; but nearly all aspects of an education system are included. To quote Hall (2000):

OBE is a comprehensive system of education organisation. It covers the contractual relationship between the government and providers (and other education bodies); the organisation and management of schools; curriculum design and implementation; summative assessment of student work; staff development and appraisal; and audit and review of institutional performance. (p. 1)

Hall (2000) reports that within the New Zealand Ministry of Education, OBE was sometimes euphemistically referred to as a *tight-loose-tight* model of monitoring. The first *tight* describes the objectives and standards that must be met by the system. For schools, this meant the introduction of more highly prescribed national curriculum with eight achievement levels and hundreds of achievement objectives. The *loose* referred to the autonomy that self-management gave to schools to implement the

curriculum. The second *tight* referred to the process by which schools were accountable for meeting their objectives. This was to take place through ERO and reporting to the Boards of Trustees. All three aspects of the tight-loose-tight model had manageability implications for assessment in New Zealand primary schools.

3.3.3 The Curriculum and Assessment in Outcomes Based Education

OBE requires a direct relationship between the curriculum and assessment. This relationship has been described by Hall (2007):

OBE comprises, in its simplest form, the declaration of the intended outcomes (objectives) of an education system or operation, the design and implementation of a programme or activities aimed at achieving these outcomes, and the monitoring of the actual outcomes against the intended ones. (p. 1)

Such a model of education has been associated with a particular view of teaching that requires a particular type of curriculum. Dixon (1999) describes this approach as a *technocratic* model of teaching:

The teacher is considered to be a skilled technician, and good practice is reduced to a set of predefined skills or competencies with little or no acknowledgement given to the moral dimensions of education (Codd, 1993; Ingvarson, 1995). As Codd (ibid) has stated, learning is reduced to the observable and measurable with the curriculum designed to produce specific learning outcomes. Of prime importance, are the products or outcomes of education. (p.14.)

The New Zealand Curriculum developed in the nineties, with its tight framework of seven essential learning areas, eight levels of achievement, and hundreds of individual achievement objectives, is an example of the type of highly prescribed curriculum associated with OBE. Such a model of education also requires national or state education systems, or in the New Zealand case, individual schools, to have assessment systems that were capable of demonstrating their pupils had learnt or mastered the achievement objectives.

3.3.4 Outcomes Based Education, Standards Based Assessment, and Criterion-Referenced Assessment

There is a strong relationship between OBE and the concepts of standards-based assessment (SBA) and criterion-referenced assessment (CRA). SBA has been defined by the Ministry of Education (2007b) as:

A standards based assessment assesses learner achievement in relation to set standards to show what students know and can do. Standards are written statements that can describe what a child has to know and be able to do in order to be awarded a unit or achievement standard and therefore gain credits towards the NCEA (or other qualification). A student's achievement is benchmarked to compare it to an expected level rather than other students' achievement (norm-based assessment).

In a less formal manner, Elley (2005) describes SBA in the following way:

What do we mean by standards-based assessment? To put it briefly, it refers to a kind of assessment where a student's achievement is judged by whether it meets a pre-ordained standard, or level of performance. Mary can or cannot recite all the letters of the alphabet. Jack can or cannot type at 30 wpm. I can or cannot run a marathon in less than 3 hours. Simon can or cannot ride a bike. The activity to be performed is clear to everyone, and most observers can agree on whether it has been reached or not. We know what the student "knows or can do."

(para. 4)

As such, SBA is closely related to criterion-referenced assessment (CRA) and the terms are often used interchangeably. Linn and Gronlund (2000) define CRA as:

A test or other type of assessment designed to provide a measure of performance that is interpretable in terms of a clearly defined and delimited domain of learning tasks. Other terms that are less often used but have meanings similar to criterion referenced: standards-based, objective referenced, content referenced, domain referenced, and universe referenced.

(p. 42)

Because the achievement objectives of the New Zealand Curriculum did not include criteria to pin down their interpretation for assessment purposes, they could not be correctly termed "standards". This point will be discussed further in Section 4.2.4.

3.3.5 Implementation in New Zealand

OBE presented fundamental changes for all elements of New Zealand education. The first major change was at the level of governance with the establishment in 1989 of self-managing schools through the Board of Trustee structure. Under this structure, elected parents' representatives would govern the school; and each school was to develop its own set of policies and procedures that met the needs of that particular community (Lange, 1988). In the view of many writers, implicit in this approach was a market model where schools were to compete with each other for students (Dixon, 1999; Elley, 2004; Gordon, 1997; Gordon & Whitty, 1997; Lauder, Hughes, Watson, Waslander, Thrupp, & Duplis, 1999; O'Neil, Clark & Openshaw, 2004; Sullivan 1993, 2002; Thrupp, Harold, Mansell, & Hawksworth, 2000). A rationale behind these reforms was that a more competitive environment would lead to improved schools. A further assumption was that differences in school communities in terms of geography, school structure, size and composition of the student population were more significant than the things that schools had in common. Both these assumptions led to a great deal of duplication of policies and procedures (ERO, 1995a).

As stated above, the change in the governance and organisation of schools was followed by the introduction, during the period 1992 to 2000, of the New Zealand Curriculum with the NZCF (MoE, 1993) and seven curriculum statements. In line with the OBE model, the 1993 Curriculum saw replacement of the very general and loosely structured previous syllabi with a very tight framework of seven essential learning areas, eight levels of achievement, and close to 1000 achievement objectives.

In 2001, the approach was further developed with the introduction of the Education Standards Act, which required each school to set and assess annual targets that were sent to the MoE.

3.3.6 Criticism of OBE in New Zealand

Elley (1993, 1996, 2004) has been a long-standing critic of the OBE curriculum reforms in New Zealand. His 2004 description of the rationale behind its introduction is particularly direct:

As stated in the Ministry's outline of the new curriculum, the Government's hope was that the curriculum documents would "specify clear objectives" and

identify for “boards of trustees, teachers, students, parents, and the wider community, a progression of desirable standards of learning against which students’ progress can be assessed” (Ministry of Education, 1993). This I believe was the hallmark of the new curricula – that the progression of outcomes in each subject would be clearly specified. They would be spelled out in the form of achievement objectives to be attained at each successive level, in each learning area, and students’ progress could then be assessed in relation to those levels. Then achievement levels would raise across the board and the education dollar be better spent. That, at least, was the rhetoric. The new model was deemed suitable for all subjects, and provided the main thrust of the Minister’s justification for a complete rewriting of all areas of the New Zealand curricula. (p. 91)

The detailed criticisms of Elley (1993, 2004) and others relating to the structure of the curriculum and the implementation of OBE in New Zealand will be covered later in the chapter.

Given the criticisms stated by the authors above, it is appropriate to consider whether there were any positive aspects in the move to OBE. In their largely critical summary of the reforms, Thrupp et al. (2000) noted four positive impacts of the change as mentioned by some teachers:

1. There was greater clarity about what should be taught, how it should be assessed and the way it should be reported to parents. Renwick and Gray (1995) also reported two teachers who were positive about changes in assessment procedures for this reason.
2. There was greater linkage between the curriculum documents, planning, resources, teaching and assessment. This point was also made by a number of teachers in Dixon’s (1999) study.
3. There were more school-wide structures for curriculum coverage and assessment than in the past.
4. There was more individualisation. Thrupp et al. reported many responses where teachers felt that they were more aware and focused on individual learning needs in their planning of teaching and assessment. This was also supported by 68% of the teachers that Wylie (1999) surveyed.

3.4 Four Major Assessment Challenges

This section briefly introduces four major assessment challenges within the New Zealand reform context. Each of these is then discussed in more detail in the following four sections.

The mathematics curriculum statement (MoE, 1992) and the NZCF (MoE, 1993a) presented New Zealand teachers with the first statements on the curriculum reforms. While the New Zealand documents were similar in structure to the earlier British reforms (Willis, 1992; Elley, 1993) they lacked the highly prescribed assessment systems based on SATs activities. New Zealand schools were expected to assess students but were given no tools to assess learning against the new curriculum. While the various documents often promoted progressive views on assessment, the fact that not enough direction and support was provided led to a challenging environment for schools in terms of the development of assessment systems.

A second major challenge that schools faced in the development of effective and manageable assessment systems was the structure and language of the curriculum documents and the requirement to assess against the levels of the curriculum in the seven essential learning areas with their numerous achievement objectives.

The third major challenge associated with the education reforms was the requirement to assess all areas of the curriculum equally. As well as differing from the British model in terms of not having a prescribed system of assessment, the publication of the NAGs (MoE, 1993c) stated that in New Zealand all seven learning areas were to have the same status in terms of assessment. New Zealand schools were required to assess and maintain individual records of all students in all learning areas. This differed from the British model where a distinction was made between what were termed the core subjects of numeracy, literacy and science and other areas of learning titled foundation subjects. This requirement to assess individual progress in all areas of the curriculum also presented schools with major challenges.

A key feature of the reforms was the establishment of ERO as an external audit agency. A significant change was that ERO reports were available to the public and media. In the new competitive environment, schools were very strongly motivated to obtain positive reports ERO reports, and thus the Review Office's views on education came to dominate practice (Dixon, 1999; French, 2000; Knight, 2000).

3.5 The Lack of Direction Given to Schools in Assessment

As stated above, MiNZC (MoE, 1992) gave New Zealand teachers and school leaders the first indication of what the new structures would mean in terms of teaching and assessment. The statements on assessment were progressive.

Assessment should focus on what students know and can do, and on how they think about mathematics. It should involve a broad range of tasks and problems and require the application of a number of mathematical ideas. Skills assessed should include the ability to communicate findings, to present an argument and to exploit an intuitive approach to a problem.

Assessment should as far as possible, be integral to the normal teaching and learning. Continuing assessment as part of the teaching and learning programme increases the range and quality of assessment which can be carried out for good diagnosis, and avoids the artificial intrusion on learning and teaching time which is associated with separate assessment sessions. Assessment should involve multiple techniques including written, oral and demonstrative formats. Grouping and team activities should also be assessed. (p. 15)

The document also suggested strongly to teachers that more was required than standard paper and pencil tests.

Traditionally, assessment in mathematics has been focused on quite a narrow range of procedures. Procedures such as pencil and paper tests of algorithms skills do not always reveal students' difficulties nor do they allow assessment of the full range of student achievements. (p. 18)

Included in MiNZC was a direction toward performance assessment and emphasis was given to diagnostic assessment. The document provided sample assessment activities:

This curriculum statement provides, at each level in each strand, examples of activities in which teachers might engage students to assess their current level of achievement. An assessment programme modelled on these examples will help teachers to plan the next stage of learning for students.

The activities illustrate assessment techniques which are not disruptive to normal classroom activities. They could be carried out as an integral part of the teaching programme rather than at times specifically set aside for "tests". Assessment and evaluation strategies of this kind require teachers to make systematic observations of students at work, and to record their observations carefully. (p. 18)

However, the examples in *MinZC* (MoE, 1992) would need considerable development in order to be used in a classroom or school programme. Furthermore, this document stated that teachers and schools were expected to develop their own assessment activities. “The few assessment activities suggested in the statement were exemplars which teachers could imitate in developing their own assessment programme” (p. 19).

Thus the new curriculum was published without any assessment tools with which to assess these new structures. The suggestion that the individual teachers and schools should develop their own assessment tools did not take account of the demands on teachers (ERO, 1995a; 2001). Some of the strongest criticism of the idea, albeit in hindsight, of schools developing their own assessment came from ERO:

Half the schools in New Zealand have fewer than 150 students. They have a small staff and teachers have to commonly teach several levels at once. It is clearly beyond them to design assessment tools as well, and unreasonable to expect that of them. (Whalley, cited in French, 2000, p. 20)

Where assessment is covered in *MinZC* (MoE, 1992), the focus is on assessment to enhance learning. There was no coverage in this publication given to assessment for accountability. However, the gazetting of the NAGs in 1993 made it a requirement that schools developed their own systems for assessment for accountability purposes. The extent to which schools were meeting the requirements of the NAGs became a focus of ERO reviews (Dixon, 1999). In this way, schools were being asked to develop methods to assess this new curriculum in the absence of tools designed for this purpose and from a different perspective than was covered in the Curriculum Statements themselves.

The NZCF (MoE, 1993a) published the year after *MinZC* was designed to “provide a coherent and integrated structure for the school curriculum” (p. 27). The approach to assessment was similar to the Mathematics curriculum statement with a strong focus on diagnostic and performance assessment and very little on assessment for accountability.

Assessment is an integral part of the curriculum. The New Zealand Curriculum builds on the close relationship between learning and assessment. It provides clear learning outcomes against which students’ progress can be measured.

Assessment in the New Zealand Curriculum is carried out for a number of purposes. The primary purpose of school-based assessment is to improve students' learning and quality of learning programmes. Other purposes of assessment include providing feedback to parents and students, awarding qualifications at the senior secondary school level and monitoring overall national education standards. Assessment also identifies learning needs so resources can be effectively targeted.

School-based Assessment

Assessment of individual students' progress is essentially diagnostic. Such assessment is integral to the learning and teaching programme. Its purpose is to improve teaching and learning by diagnosing learning strengths and weaknesses, measuring students' progress against defined achievement objectives, and reviewing the effectiveness of teaching programmes. The information which teachers record from these assessments enables clear profiles of individual student's achievements to be built. These profiles are used to inform teachers about each student's learning and development and to provide the basis for feedback to students and parents. (p. 24)

As with *MiNZC* (MoE, 1992), the NZCF (MoE, 1993a) gave very little attention to assessment for accountability purposes. The NZCF introduced *The Essential Skills* in the New Zealand school system. The eight essential skills were: Communication Skills, Numeracy Skills, Information Skills, Problem Solving Skills, Self Management and Competitive Skills, Social and Co-operative Skills, Physical Skills and Work and Study Skills. These were broken down into a further 58 sub-skills. The assessment of the essential skills proved problematic for schools. A number of them were largely straightforward:

Communication Skills

Communicate competently and confidently by listening, speaking, reading and writing, and by using other forms of communication where appropriate

Information Skills

Identify, locate, gather, store, retrieve, and process information from a range of sources. (p. 18)

Others were more challenging, some because of the need to assess very high level skills:

- Problem Solving Skills
- Think critically, creatively, reflectively and logically
- Exercise imagination, initiative, and flexibility
- Identify, describe, and redefine a problem

- Analyse a problem from a variety of different perspectives. (p. 19)

Others were either far too subjective to measure or involved life-long issues that would be inappropriate for teachers to assess, or both:

Self-Management and Competitive Skills

Show initiative, commitment, perseverance, courage and enterprise

Develop constructive approaches to challenge and change, stress and conflict, competition, and success and failure. (p. 19)

The developers of the NZCF acknowledge this issue when they state in a footnote:

It is acknowledged that some qualities and values such as rangimarie (tolerance) and integrity are difficult to measure. Nevertheless, the development of these attributes is vital to a student's learning and will be assessed as part of the school curriculum. (p. 21)

When evaluating schools' assessment systems, ERO mostly did not distinguish between those essential skills that were within a school's capacity to measure and the latter, where those making the judgements would need extreme wisdom to "divine" a student's ability on some skills, or to spend a great deal of time in and out of school to judge if they could, for instance: "Participate appropriately in a range of social and cultural settings" (p. 24). While it is a widely-held view that education should be about high ideals and life-long development, it can be argued that there is a serious limitation on how realistically these goals can be assessed within normal classroom contexts.

In 1994, the MoE attempted to further clarify assessment issues with the publication of *Assessment Policy to Practice* (MoE, 1994b). The MoE still held the view that teachers and schools should develop their own assessment tools.

Although the examples in the national curriculum statements provide a source of ideas, the sets of tasks described are neither exhaustive nor definitive. Teachers will also need to locate and devise their own assessment tasks for diagnostic, monitoring, and review purposes. (p. 5)

Associated with this belief, the MoE approach was still based on the idea of each school developing their own policies and procedures. "The purpose of the policy is to translate national requirements into practice, taking into account the school's own community, strengths, resources and special circumstances" (p. 12). Unlike the

previous MiNZC (MoE, 1992) and NZCF, the document *Assessment Policy to Practice* (MoE, 1994b) acknowledges the role of summative assessment in reporting to stakeholders. The publication even included a short statement on the manageability of assessment. “Assessment must be manageable in the classroom. There is an inevitable tension between validity, reliability, and manageability. Compromises are often necessary and should be acknowledged and allowed for in evaluating the results of assessment” (p. 25).

In response to the lack of assessment tools to assess the new curriculum, schools often reverted to using assessment tools that they were familiar with but which did not match the new curriculum (ERO, 2001; French, 2000). ERO also recognised this lack of guidance and the provision of assessment tools as a problem (ERO, 1995b). However, their solution was more radical; a campaign to introduce national testing (ERO, 1998).

3.6 The Structure of the Curriculum

While the lack of specific guidance on assessment and the absence of assessment tools for the new curriculum provided challenges to schools, it could be argued that the structure of the curriculum provided an even bigger challenge. MiNZC (MoE, 1992) introduced New Zealand teachers to the structure of eight levels of the curriculum and the corresponding achievement objectives that were going to be used in the other seven curriculum documents.

Critics challenged the structure of the curriculum, the number of achievement objectives and the lack of precision in language used for both the levels and the objectives. Elley (1993, 2004) pointed out there was no rationale for dividing the New Zealand curriculum into eight levels or the British into ten. There was no research on children’s learning to support the level structure. He also stated that the level structure was unhelpful for teachers and they often had to create additional structures such as low 3, middle 3, and high 3, to implement them in classrooms. Elley (2004) concedes that a hierarchical level-based curriculum may be helpful in such things as psychomotor skills and where there are clearly definable competencies such as some elements of mathematics learning. Such an approach, according to Elley (2004), makes no sense in many areas of learning.

Many researchers have tried to identify the progressions in reading and spelling, scientific thinking and intellectual development and moral maturity, to name but a few, with only modest success. Rarely do researchers agree, and their stages or levels are fraught with reversals and mid-classification. The problem is that students' knowledge growth in most topics in the curriculum is individual and idiosyncratic. Their knowledge develops in unique networks of partially-mastered concepts and multiples of particulars. Each student organizes his or her cognitive systems in individual ways, depending on their unique experience. Rarely are their concepts mastered in an all or nothing fashion. Typically their understanding expands in irregular spurts, sidetracks, inconsistencies, and new insights. So students master Level 5 objectives before Level 4 in some specific areas while lagging at Level 2 in others of the same strand. (p. 94)

A further problem with the structure of the curriculum lies in the language of the achievement objectives. As Elley (1993) stated:

The futility of trying to impose a tight level structure on a subject like English was well illustrated in the actual wording of the achievement objectives for successive levels. How can teachers decide what level pupils have reached in "Close Reading" when the level structures are worded like this?

Level 1: Respond to language and meanings in texts.

Level 2: Respond to language, full meanings and ideas in different texts, relating them to personal experiences.

Level 3: Discuss language meanings and ideas in a range of texts, relating their understanding to personal experiences and other texts. (Ministry of Education, 1994).

In my view, these vague statements do little to capture the key differences in close reading between 5 year-olds at Level 1 and 10 year-olds at Level 3.

Looking more closely at the National Curriculum statements cited above, I suspect that a logician would have a field day with some of the progressions. We could ask, for instance, what is the difference between meanings and progressions? What is the force of the change from "in texts" to "in different texts" or "in a range of texts"? (pp. 95-96)

ERO (1995, 2001), Le Metais (2002) and Donnelly (2007) also expressed concerns about the language of the achievement objectives. However, their arguments differed from Elley (2004) in that they should not be seen as a criticism of the curriculum model itself, rather a view that a lack of precision in the language of the achievement objectives meant that they were inappropriate for the successful implementation of such an approach. As ERO stated in their 1995 publication *Assessing Student Achievement*:

Many of the achievement objectives in the National Curriculum Statements are also not sufficiently specific for a judgment to be made about their achievement. For example, a Level 3 achievement objective in Mathematics in the New Zealand Curriculum is: “perform measurement tasks using a range of units and scales”. In this objective neither the “tasks” nor “range” is sufficiently defined for an inference to be made about what a student is actually required to do to achieve this objective. It will therefore be difficult to know whether a student has met this objective. (pp. 4-5)

In another publication, ERO (2001) stated “achievement objectives are often too broadly stated to indicate what can be reasonably be expected of students. Most curriculum statements lack clarity about how the achievement objectives are differentiated between levels, and are vague in terms of the content of the levels” (p. 2).

Le Metais (2002), in a MoE funded evaluation of the New Zealand Curriculum, expressed concerns about the achievement objectives:

The achievement objectives are the most contentious and arguably, the weakest, element of the curriculum documents. They are perceived, and act, as a constraint on teaching and learning, but they do not lend themselves to reliable assessment and meaningful recording of performance. (p. 16)

Donnelly (2007) stated that the achievement objectives should not be confused with standards:

The use of words such as levels in National Curriculum Statements, the New Zealand Curriculum Framework and the National Qualifications Frameworks implies a consistency of standards. This is not, in fact, the case.

Standards of achievement are to be found more in the rhetoric than the reality. The National Curriculum Statements do not define standards but describe achievement objectives for a number of levels.

The sheer number of achievement objectives and the variation in their specificity means that teachers individually and nationally will have difficulty in reaching consistent judgments about when students reach certain levels of achievement. (p. 32)

A criticism of the achievement statements introduced with the curriculum is that the approach is essentially behaviourist in its nature in that it reduces learning to closed and prescribed units of study. Such an approach is at odds with much of the teaching–learning context in our schools, which is based on constructivist and socio-cultural views of learning (Dixon, 1999; Knight, 2000).

There is not a great deal of evidence available that records what schools were actually doing during the early period of the reforms in response to the requirement to develop assessment based on the levels of the curriculum and the achievement objectives. However, the written advice given to teachers and schools by those working as professional development advisers does provide some information on school practices. Writing during the period 1995 to 1997, Baker and Lorrigan (Baker, 1997; Baker & Lorrigan 1995, 1996) suggested a number of assessment developments for schools:

- Establishment of an assessment committee
- A review of the school's existing assessment policy
- Consultation with the community
- Develop a school wide assessment statement
- Develop an assessment plan for each specific learning area
- Develop achievement statements in each specific learning area
- Develop an Analysis/Log/Mark Tracking book for each teacher
- Develop student portfolios
- Develop school assessment portfolios for each learning area. That is a portfolio that explicitly states the school's standards and exemplars for each learning area
- Develop record of achievement/Individual student profiles. This is the translation of class data onto individual student files to fulfil legal recording and reporting obligations

(Baker & Lorrigan, 1995, p. 4)

The ambitiously titled *Easy Assessment in Social Studies* (1999) by two Wellington school advisers, McMaster and Bonallack, also illustrates the advice schools were being given during this period. They suggested that for each social studies unit, one knowledge outcome and one process skill should be assessed and recorded. They also suggested an annotated sample of each child's work be collected and placed in the individual pupil folder. At the end of the unit, McMaster and Bolland suggested the gathering of summative data. They also suggested a three point assessment against the achievement objective:

3 = Can do it well;

2 = Can do it;

1 = Can't do it yet

(p. 45)

Such advice when applied to all curriculum areas and the numerous achievement objectives they contained was highly questionable in terms of manageability and utility. It was also questionable how such processes would lead to improved student achievement.

In their mathematics teaching resource which they sold to other schools, Wellsford School (2001) suggested a similar three-point approach for assessment: “CA for Competent Achievement, SA for Satisfactory Achievement and HD for Having Difficulty” (p. 2). As stated earlier, this type of three-point assessment became common in many New Zealand schools (Hill, 1999; 2000; Dixon, 1999; Knight, 2000) and was seen as a response to government requirements for accountability and had little if any formative or diagnostic purposes. The quote below from a teacher in Hill’s (1999) study summarises the approach:

In my roll book I’ll have the learning outcomes...with a tick, plusses and minuses...I try to cut it down and keep it to yes, no, sometimes that sort of thing.

I am teaching to level 3, then that’s what I am assessing unless a child is below that and then I probably write level 2 next to their name [I know they are below] if they haven’t got a tick with a tail or a cross through it.

So as the children are working, the assessment is done, not based on tests alone...and so we mark them off with a dot, slash or cross. (p. 181)

Knight (2000), whose work brought her into contact with many schools, questioned the utility of such activity:

In my capacity as a facilitator for the MoE Assessment for Better Learning professional development programmes, I have viewed many teachers’ planning and assessment folders which contain hundreds of carefully listed specific learning outcomes. These are systematically ticked or dotted when an outcome is achieved or not achieved. When asked about the value of this practice in enhancing student learning, teachers struggle to comment positively! (p. 42)

Dixon (1999) concluded from her study of the assessment practice of Auckland teachers that:

The use of assessment methods informed by behaviourist assumptions about teaching and learning (for example checklists) had been used frequently by teachers in their initial attempts to assess children’s curriculum achievements.

These were viewed quite critically by teachers as they questioned the amount, relevancy, dependability and usefulness of such data. (p. 161).

In the area of reporting to parents, the MoE's 1994 Assessment Policy to Practice publication gave directions that in hindsight would have been hard to justify.

As part of the report to parents, there should be included a statement of the level at which a student is working in each curriculum area. As each student may be working at different levels within each curriculum area or strand of each curriculum area, this information should be given for each strand of each curriculum area specified in the national curriculum statements, for example, in mathematics, separately on Mathematical Processes, Number, Measurement, Geometry, Algebra and Statistics. The level(s) the student is working at should be determined by the teachers' professional judgement based on sound evidence. (p. 38)

Planning and Assessment in English (MoE, 1997b) gave more detailed guidance to teachers on the implementation of *English in the New Zealand Curriculum* (MoE, 1994a) and included information on reporting to parents. This publication included the concept of a *best fit* approach to reporting. It was suggested that teachers needed to sample students' achievement from a number of strands and levels of the curriculum before deciding where the best fit was for each individual. This suggestion addressed concerns that children progress at different stages in different strands of curriculum. Whether using a best fit approach or not, as stated above, many teachers found assessing against the levels to be a burden and was no use to them in terms of real assessment or teaching (Dixon, 1999; Hill, 2000; Knight, 2000). While the approach was also based on the idea that students' achievement would be reported to parents in terms of the level of achievement, Peddie (2000) concluded that this was not a common practice. Whatever the logic or merits of the approach, *assessment against the achievement objectives* and *using a best fit model*, or more commonly the lack of it, became a common phrase in ERO reports for many years:

Summative assessment judgements are not explicitly linked to curriculum levels. Therefore, the school cannot report to parents or the board how well students are progressing according to the levels of the curriculum.

(St Josephs Upper Hutt, ERO Report, February 2004)

Thus in terms of the manageability and utility of assessment, the NZCF presented problems. As stated above, there were a high number of achievement objectives to be assessed. In addition, the level structure was debatable and the language of the numerous achievement objectives lacked clarity (Elley, 1993, 2004; Donnelly, 2006;

ERO, 2001; Le Metias, 2002). The rushed development of the curriculum meant that it was not until 10 years later with the development of asTTle that the levels of the curriculum were subject to a level of professional scrutiny that was required for such major documents.

3.6.1 asTTle and the Levels of the Curriculum

The challenge presented by attempting to assess against the levels of the curriculum was illustrated in the development of asTTle. The NZCF (MoE, 1993a) and each Curriculum Statement provided a table that gave some guidance between the match for year level and curriculum level. A copy of this table is shown below as Figure 3.1. For example, Level 4 was located as being somewhere in between year 8 and year 9. Included in the brief given to the developers of asTTle, was a requirement to develop a test that measured reading, writing and mathematics achievement in the context of the prescribed levels of the New Zealand curriculum. As part of the development of asTTle, a curriculum mapping exercise was undertaken. This lengthy process involved experienced teachers, and literacy and numeracy experts unpacking achievement objectives and matching them against various test questions (Ell, 2001; Thomas, Tagg, Holton, & Brown, 2002). In order for the achievement to match more accurately and finely each level of the curriculum, levels were further broken down into sub-levels described as *basic*, *proficient* and *advanced* (Ell, 2001; Hattie, Brown, Keegan, Irving, McKay & Sutherland, 2003; Meagher-Lundberg & Brown 2000). Thus, instead of the original four levels of the Curriculum that asTTle was originally designed to measure (Level 2 to Level 5) these benchmarks were expanded to twelve:

Level Two Basic	Level Two Proficient	Level Two Advanced
Level Three Basic	Level Three Proficient	Level Three Advanced
Level Four Basic	Level Four Proficient	Level Four Advanced
Level Five Basic	Level Five Proficient	Level Five Advanced

The final asTTle tests (MoE, 2007) reported the year group mean as much as two levels lower than the guidelines given in the original Curriculum Statements regarding the match between year level and achievement level. This is a complex, but important point. Thus, according to asTTle Reading, the average year 8 student is operating at Level 3 basic – considerably lower than the guidance that teachers were given when *English in the New Zealand Curriculum* was published in 1994. The effect of

asTTle's considerable work on the levels has, if we take their views as correct, dramatically lowered equivalence between the levels of curriculum achievement and where the average child was thought to be operating.

Figure 3.1: Curriculum Levels and Year Levels in the New Zealand Curriculum

(Mathematics in the New Zealand Curriculum MoE, 1992, p. 17)



Furthermore, asTTle's view on curriculum levels as stated in their technical reports (Ell, 2001; Hattie et al., 2003; Meagher-Lundberg & Brown, 2000) and the tests themselves (MoE, 2007) brings into question ERO's long-standing criticism that some schools were not operating a best fit assessment system for levels of the curriculum. If the curriculum developers had it seriously wrong in the first place, why should individual schools have been expected to get it right? Given that asTTle has spent thousands of hours to come to the conclusion that the levels of achievement were inaccurate in the original curriculum documents for mathematics and English, the safe assumption is perhaps to treat all curriculum areas as being potentially inaccurate in relation to the match between year of schooling and curriculum level.

The above discussion illustrates that while levels of the New Zealand curriculum might have provided appropriate guidance for teachers for planning purposes, they were inappropriate for assessment. This makes the brief for the development of asTTle extremely challenging as it is very difficult to base an assessment system on something that itself is lacking in well articulated logic, consistency or a well-

developed understanding of how children learn (Elley, 1993, 2004). It would appear that far more work has been put into the asTTle project on measuring the levels of the curriculum than was ever put into their initial establishment. The asTTle Project has certainly left a far more substantial paper trail (Ell, 2001; Hattie et al., 2003; Meagher-Lundberg & Brown, 2000; MoE, 2007) than was the case with the original development of the curriculum documents.

3.7 The Lack of Distinction between the Core Curriculum and the Other Areas of Learning

The NAGs (MoE, 1993c), which took effect in April 1993, detailed the requirements for BOTs.

The new guidelines will form a major part of the contractual arrangements boards of trustees have with the Crown. As such they will be a significant part of the basis for audits and reviews conducted by the Education Review Office.

In the area of assessment and reporting they stated:

Through the principal and staff, boards of trustees are required to:

- I. Implement learning programmes based on the underlying principles, stated essential learning areas and skills, and the national achievement objectives;
- II. Monitor student learning progress against the national achievement objectives; and
- III. Assess student achievement, maintain individual records, and report on student progress

(Education Gazette, 30 April, 1993, p. 3)

The Review Office took these statements literally. In carrying out their audit function they reviewed schools' assessment systems in all areas of the curriculum, all achievement objectives, and all essential learning areas (ERO, 1994). That individual schools could be expected to develop both assessments systems and tools that carried out these functions, while at the same implementing a radically new curriculum was recognized later as demanding too much (Fancy, 2004b; Knight, 2000; Renwick & Gray, 1995; Wylie, 1997, 1999).

Schools' responses to the requirement to assess all areas of the curriculum and maintain individual records led to the keeping of students' assessment information in variety of formats and names, which included *student folders*, *profiles*, *clearfiles* and in some case *portfolios*. The use of the word *portfolios* was to some extent problematic, as many of what were termed *portfolios* would not comply with how advocates of performance assessment understood the term. This issue will be discussed in more detail later in the chapter.

In the same way as schools were guided in adopting the three-step achievement objectives, school advisers were advocating the use of student folders. To quote Baker and Lorrigan (1995) again, schools were advised to:

- Develop student portfolios
- Develop school assessment portfolios for each learning area. That is a portfolio that explicitly states the school's standards and exemplars for each learning area
- Develop record of achievement/Individual student profiles This is the translation of class data onto individual student files to fulfil legal recording and reporting obligations

(p. 4)

While the evidence is not extensive, it appears that the utility of these folders was questioned by teachers: "it's done for the next teacher and I know she won't use the information even if she bothers to read them" (Hill, 2000 p. 173). The same attitude was also expressed by teachers in Symes and Timperley (2003) and Knight (2000). Williams (1999), in her study of assessment procedures at a Marlborough school, captures some of the same concerns with what were termed *clearfiles*.

The clearfiles also contained other formal tests, assessments and samples collected by individual teachers. However no criteria were set for samples and formal tests contained a raw score but rarely any detailed analysis. Once again no school-wide system guidelines seemed to apply and at the beginning of each year as files moved on with the child each teacher would delete or add to the file as they saw fit. (p. 14)

In line with the direction of the NAGs (MoE 1993c), a common phrase in early ERO reports was the *lack of a balanced curriculum* and the implication that too much emphasis was being given in teaching and assessment of literacy and numeracy in comparison with the other essential learning areas (ERO, 1995a).

The Literacy Task Force (1999a), a group of teachers and academics formed by the government to address perceived concerns about literacy levels, expressed concern about the amount of assessment teachers were required to carry out. Their concern focused on the NAG requirement to teach a balanced curriculum and, by implication, how ERO was interpreting that requirement in school reviews.

The taskforce was concerned that the way in which the NAGs requirements “to monitor student progress against the national objectives” and to “assess student achievement, maintain individual records, and report on student progress” have been interpreted and implemented is not reasonable. The taskforce was aware of teachers being required to gather large amounts of detailed data to record the progress of individual children across the curriculum at the expense of quality instruction time. It is difficult for teachers to maintain an emphasis on literacy and numeracy in these circumstances. The taskforce believes that all the essential learning areas are important particularly for those children in the target groups because of the rich experiences they provide, for example, in art and physical education, but the workload associated with monitoring in this way often intrudes on quality teaching and time. (p. 15)

The Taskforce’s Report was influential in having the MoE modify the NAGs in 1999 to give greater priority to literacy and numeracy in years 1 to 4 (MoE, 2000b). However, it needs to be noted that this modification only applied to years 1 to 4. Expectation for the assessment of older children in primary school had not changed.

3.8 The Impact of the Education Review Office

The Education Review Office had a troubled birth. It was established in October 1989 and carried out its first school review in July 1990 (French, 2000). However, before its first review had been carried out, the Office had been reviewed and its funding cut. In October 1990 ERO’s first CEO, Maurice Ginaotti, resigned a year after the Office’s establishment. By 1992, there had been two rounds of redundancies for newly appointed ERO staff. However, by 1992-1993 the Review Office had, under the direction of its new CEO, Judith Aiken, implemented training of its staff and clarified for the Office itself (if not for schools and other education institutions) its roles and procedures. From 1993 onwards, it began to have an increasingly strong influence on education in New Zealand. In the words of its staff, ERO reports beyond this period became “sharper, clearer and more direct” (French, p. 24). French also notes that under Aiken’s leadership, ERO made skilled and deliberate use of the media.

The influence of ERO was strengthened because the New Zealand reforms differed from those of other western countries in that a system of national testing was not introduced. This absence of a national testing regime meant that ERO reports became the main form of high stakes assessment in the New Zealand primary school education system (Fancy, 2004a). In 1999 Dixon researched assessment practices by questionnaire and interviews in four Auckland schools. She stated:

However, it was the Education Review Office (ERO) that was considered to be the agency which enforced these increased demands and requirements. Writers such as Clark (1998), O'Neil (1998) and Robertson et al. (1997) have argued that ERO in its monopoly position as the ultimate arbiter of acceptable practice, have been a driving force in bringing about changes in school and teacher practice, particularly in the area of assessment. Many of the teachers in this study supported this view:

“ERO forced changes on us.”

“Because ERO expected them.”

“It is required, ERO look for these.”

“...because of the pressure of ERO,

I understand more individual assessment is done to meet the requirements of ERO.”

“I feel assessment is being driven by ERO.” (p. 67)

From her research, Dixon concludes that during the period 1993 to 1999, ERO had a major influence on assessment practice in New Zealand primary schools.

Within the context of this study professional attempts to initiate change had been tempered by the Education Review Office. This was true for all schools involved in the study as it appeared the prospect of a review, either in a given year or at some unknown time in the future, loomed large in teachers' minds. ERO's requirement that school performance be reported in relation to outputs, contractual compliance and the attainment of specific outcomes had resulted in a professional dilemma for teachers as they participated in certain assessment practices which were at odds with their personal beliefs about the efficacy of such practice. In an attempt to meet external accountability requirements, teachers engaged in practices, many of which they themselves did not consider relevant to the improvement of teaching and learning, but which they perceived as important if their school was to be judged, by ERO and the wider community, as providing quality education. (p. 42).

However, the Review Office was not unaware of these issues:

“They [primary schools] do a lot of assessment, and when they found out we were asking about it they did even more”; and “Teachers were desperate to get some information for us”; and “they would go back to old fashioned tests that didn’t match the curriculum” (Frances Salt, cited in French, 2000, p. 24).

Robertson, Thrupp, Dale, Vaughan, and Jacka (1997) carried out a review of ERO for the secondary teachers union, the Post Primary Teachers Association (PPTA). This review concluded that the focus of ERO reviews were on compliance factors (managerial and administrative) rather than education processes and encouraged a preoccupation with paperwork and a failure to look at practice. They also suggested that ERO “create a production regime in schools which leads to the manufacture of artefacts and ritualistic displays in order to make work visible” (p. 6). This report did note that “many respondents were largely or completely satisfied with ERO officers’ practice in the school but with teachers less satisfied than principals” (p. 6). Such views were supported by Snook, Adams, Adams, Clark, Codd, Collins, Harker, O’Neil, & Pearce (1999).

While teachers and principals often blamed ERO for assessment in schools becoming unwieldy and increasingly dominating teaching during the period of the mid-nineties, the Office’s first publication on assessment, *Assessing Student Achievement* (ERO, 1995a), was a perceptive criticism of a number of faults with the MoE’s policies. ERO pointed out that with more than 200 achievement objectives in one curriculum statement alone (Mathematics) in Levels 1 to 4, there was a risk of assessment becoming unmanageable. The Office noted that while a national transition point assessment was being developed, the items would not cover all learning areas. They also criticised the lack of assessment tools available to assess the curriculum. The publication also made the point that the majority of the MoE’s publications focused on the assessment of individuals and gave little attention to the type of summative assessment that schools were required to do in terms of accountability. The issue of each school having to reinvent the wheel was also raised.

It is inefficient and unrealistic to expect each of the 2700 schools nationally to devise its own tools and practices. Schools can invest significant time and energy in developing assessment policies and procedures which result in each school “re-inventing the wheel.” (p. 34)

3.8.1 ERO and National Testing

The last years of the 1990s were concerned with the issue of national testing. The debate was initiated by ERO (1998). The Office made a number of claims in favour of national testing. They stated that there was little dependable information about students' achievement. Where school-based assessment was taking place against the achievement objectives, there was no way of ensuring that there was consistency between schools. The Office also expressed concerns about the effectiveness of all schools having to develop their own systems and assessment tasks. They argued that the existing system did not provide adequate information for accountability purposes for parents, boards of trustees and the Crown. Associated with this argument was the view that without some form of standardised national testing, ERO was incapable of performing the auditing function it was charged with. ERO also stated that national testing would provide schools with better quality information with which to improve teaching and learning.

The *Assessment for Success in Primary Schools Green Paper* (MoE, 1998), which echoed ERO's concerns, contained the MoE's draft proposals on national testing. In addition to national tests there were to be a number of other assessment initiatives developed (Phillips, 1998).

However, the introduction of national testing had long been a concern of the teachers' union, the New Zealand Educational Institute (NZEI), and the New Zealand Principals' Federation. Their concerns were that schools would be ranked in *league tables* and the focus of teachers would be on teaching to the test (Dominion, 8 May, 1988). Observations and horror stories from the UK and the USA dominated many comments. While their concerns were very understandable they came out as defending the existing system.

Information for better learning (MoE, 1999c) was the MoE response to submissions to the Green Paper and contained their policy directions and initiatives for the future. There was to be development of new diagnostic tools especially in literacy, numeracy and Maori medium education (Phillips, 1998). This development was to be accompanied by a stock-take of existing tools (MoE, 2002a). Exemplars were to be

developed in English and Maori for Levels 1-5 of the curriculum in each of the essential learning areas. The contract for the ARBs was to be extended, with the ongoing development of curriculum-based assessment activities in mathematics, English and science. This was to be accompanied by an independent evaluation of the ARBs. NEMP was to be expanded to provide more information on the achievement of Maori and Pasifika students. It was also decided to include probe studies with NEMP to explore data further. More NEMP tasks were to be released to schools.

However, the controversial decision in *Information for better learning* (MoE, 1999c) was the introduction of a pilot of externally referenced tests in literacy and numeracy. The aim of the pilot was to develop assessment tools that could:

- Monitor student learning
- Inform decision making with respect to teaching programmes
- Improve the consistency and effectiveness of reporting to parents using nationally comparable information. (p. 22)

The contract for the pilot was allocated to the University of Auckland, with Assessment Tools for Teaching and Learning (asTTle) being the final product. However, with the election of the Labour Government in 1999, the plan for national testing was, if not dropped, definitely shelved. asTTle was released to schools in 2003 but its use was optional.

3.8.2 The Changing Role of ERO

In 2000, the roles and responsibilities of the Education Review Office were reviewed by a committee chaired by former cabinet minister Stan Rodger. The report (MoE, 2000) recommended moving from a model with less focus on compliance and include a capacity to *advise and assist*. Although not dramatically changing their approach, the new century saw ERO reviewers spending more time observing in classrooms and less reviewing paper work. The Office also put increasing emphasis on formative assessment and the use of assessment information in schools.

3.9 The Response of Schools to the Assessment Requirements of the Reforms

The increased documentation as suggested by Baker and Lorrigan (1995) and McMaster and Bonallack (1999), and illustrated by Hill (2000, 2000a, 2003), had a

serious effect on the morale of teachers. Similarly, the rapid pace of curriculum change put a great deal of pressure on schools to develop systems to implement this radically new approach to the curriculum (Capper, 1991; Fancy, 2004a; Livingstone, 1994; Renwick & Gray, 1995; Willis, 1992; Wylie, 1997, 1999).

The response of schools was illustrated in a number of studies. Bridges (1992) surveyed Christchurch teachers who reported increased working hours and a pessimistic outlook as a result of the changes in practices brought about by the Tomorrow's Schools reforms. Livingstone (1994) conducted research into the workloads of 157 Wellington primary school teachers for the Wellington region NZEI. The six most stressful workload factors identified by teachers were:

- Amount of paper work required
- ERO reviews
- Implementation of the new curricula
- Development of new assessment procedures
- Number of hours teaching/ at work
- Ways in which change is being implemented

(p. 31)

Renwick and Gray (1995) and Wylie (1997, 1999) provided insight into the effects that the curriculum reforms were having on teacher practice. Renwick and Gray reported that teachers had found the pace of the curriculum reforms challenging with most reporting increased workloads. While many of the teachers in the seven case study schools were positive about the curriculum statements, assessment was an issue with a large number of teachers. There was, in Renwick and Gray's words, "uncertainty about how much should be assessed and appropriate assessment methods" (p. ix). Wylie in her conclusion from a series of five teacher, principal, trustee and parent surveys (covering 1989 to 1996) reported that teachers had greatly increased the amount of assessment they were doing. While it was reported that this increased assessment did have some benefits such as gaining a better picture of individual children (40% of respondents), concerns were raised that increased assessment was leading to less attention for individual children during class (44%), less time for planning and preparation (39%), and more work out of class hours.

3.10 The Professional Development of Teachers: Assessment for Better Learning

For assessment to be effective and manageable, it is important that teachers and school leaders have a well-developed understanding of assessment. In 1996, ERO expressed concern about New Zealand teachers' lack of professional understanding of assessment. Hill (2000), in her study of teachers' assessment practices, also expressed concern about New Zealand teachers' lack of assessment literacy.

The Assessment for Better Learning (ABeL) professional development contract began in 1993. The title later changed to Assessment to Learn (AToL). By 2004, 27% of primary schools had participated in the programme (Rose Hague, pers. comm., 18 July 2005). The programmes have aimed to develop teacher knowledge of effective assessment and to raise students' levels of achievement. It includes elements of reviews of schools' existing assessment systems, developing understanding of the role of formative assessment, effective use of assessment tools, analysis of assessment data, and improved reporting to parents and BOTs.

In 2000, Auckland University (Peddie, 2000) was contracted to evaluate the contract from the period 1995 to 1999. The report concluded that ABeL had successfully provided participating schools with effective systems which included data aggregation, school policies, better reporting to parents and BOTs, and increased understanding of formative assessment. However, as noted earlier in the chapter, Peddie identified that very few of the schools that had undertaken the professional development were reporting to parents on levels of achievement.

3.11 Planning and Reporting

In October 2001 the Education Standards Act (MoE, 2001a) was passed and included what has been termed *planning and reporting*. The changes required schools to update their charters on an annual basis and include specific targets for children to achieve each year. At the end of the year schools were to report on their achievement of the targets. These changes increased the workload on schools. The fact that each school was to develop their own targets and assessment processes meant that Government retained some commitment to the concept of self-managing schools (McMahon, 2000).

3.12 Recent Developments

The changes announced in *Information for Better Learning* (MoE, 1999c) and the *National Assessment Strategy* (Chamberlain, 2000) addressed some of the issues associated with the rushed and shallow development of the curriculum released only seven years before. With the development of The New Zealand Curriculum Exemplars, the expansion of the ARBs and externally-referenced tests, there would now be a number of tools available for schools to assess learning within the curriculum. The concept of each school developing their own assessment tools had largely been abandoned. Furthermore, there was a realisation that schools needed to do less assessment (MoE, 2000a). Other initiatives by the Ministry of Education since 2000, included in the *National Assessment Strategy*, have aimed at supporting teachers and addressing current issues in assessment. The *Assessment Community* on the TKI (<http://www.tki.org.nz/r/assessment/>) website provides comprehensive support for teachers. This includes access to the main assessment tools (ARBs, The Exemplars, and asTTle), school case studies, and professional reading. This site and the MoE publication *Curriculum Update* gives guidance to teachers on the Ministry's major assessment themes of increased formative assessment and evidence-based teaching. These changes also reflect an emphasis on improving the quality of teaching in New Zealand as a major strategy in addressing under-achievement (Alton-Lee, 2003; MoE, 2004; Thrupp, 2004, 2005).

In 2000, the Ministry of Education began reviewing the curriculum, resulting in The New Zealand Curriculum: Draft for consultation (MoE, 2006) and, after consultation, *The New Zealand Curriculum* (MOE, 2007). The new curriculum contains eight learning areas with a substantial reduction in the number of achievement levels of the curriculum. Another major change is that the Essential Skills have been replaced by five Key Competencies: *Thinking, Making Meaning, Relating to Others, Managing Self, Participating and Contributing*.

The competencies and the essential skills do not differ greatly in terms of assessment issues. While it seems reasonable that education should include a focus on learning and life-long goals, the challenge is that such goals are extremely difficult to assess validly and reliably without devoting a lot of time and resources to this goal. While it

is relatively simple for schools to demonstrate that they are providing programmes that address these skills and competencies, it is extremely challenging to assess how successful an organisation is in achieving these goals for individuals and groups of students as well as reporting these to parents, government agencies and BOTs.

At the start of this chapter, it was stated that one of the arguments put forward for the changes associated with the reforms was the view that competition between schools would lead to increased student achievement and self-managing schools would make schools more responsive to their local communities. By 2004, the Ministry of Education had retreated from this position, both in the views of its leader and its operations. This speech from Howard Fancy (2004a), Secretary of Education 1995-2007, captures the change:

Tomorrow's schools emphasised self management. I now see the "emerging paradigm" acknowledging that a successful system also needs to be characterised by networks of relationships that contribute to a system and all its constituent parts being unambiguously focused on raising achievement. I think we have learned that schools cannot be "isolated islands" in themselves but also need to be seen as archipelagos with measures of both independence and interdependence.

This view is also evident in a number of initiatives by the Ministry. These include projects by the MoE, such as the South Auckland programme titled *Picking up the Pace* (Phillips, McNaughton, & MacDonald, 1992), I CAN (Porirua) and the Taita Central Literacy Project (Mallard, 2000, 2003), which all involved clusters of schools, academics and the Ministry working together to enhance student achievement. One particular assessment feature of such initiatives is the more scientific application of assessment information to teaching in what has been termed *evidence-based teaching* and the increase in the assessment literacy of teachers involved in the projects.

3.13 International Influences

At the start of this chapter, it was argued that to fully understand developments in New Zealand assessment practice, it is necessary to examine not only local policy and initiatives but also influences from overseas on New Zealand assessment practice. This final section of the chapter examines four major developments and their influence on assessment practice in New Zealand schools: Authentic or performance assessment; formative assessment; evidence-based teaching; and international testing.

3.14 Authentic or Performance Assessment

Advocacy for what has been termed performance or authentic assessment began in earnest in the eighties and had a significant effect on policy and practice in the nineties. However, several authors (Madaus & O'Dwyer, 1999; Ward & Murray-Ward, 1999) have traced performance assessment back to 2000 BC in China when candidates for the civil service ranks had to go through a series of performance tasks. The authors also trace performance assessment to the craft guilds through the later middle ages and oral examinations in the early universities. In fact, they argue, when considered in historical terms, paper and pencil quantitative assessment has been the exception and performance measurement has dominated assessment.

Definitions and Associated Terms

There are an extensive number of terms associated with this concept, with different terms highlighting a particular emphasis. Linn and Gronlund (2000) state:

Performance assessments are also sometimes referred to as “authentic assessments” or “alternative assessments”. But the terms are not interchangeable. “Alternative assessment” highlights the contrast to traditional paper-and-pencil tests; “authentic assessment” emphasises the practical application of the task in real world settings. (p. 260).

Worthern (1993) notes that whether the approach to assessment is called direct assessment, authentic assessment, performance assessment or the title he uses, alternative assessment, they have two key features in common. They are all viewed as alternatives to multi-choice, standardised achievement tests. Secondly, these approaches in Worthen's view “all refer to direct examination of student performance on significant tasks that are relevant to the outside world” (p. 445).

McTighe (1996/1997) makes a similar point when she states: “performance tasks should call upon students to demonstrate their knowledge and skills in a manner that reflects the world outside the classroom” (p. 3).

Madaus and O'Dwyer (1993) use the three Ps term to describe performance assessment – performance, portfolios and products.

Meisels, DiPrima Bickel, Nicholson, Xue Atkins and Burnett (2001) use the term *curriculum embedded performance assessment*, which is “integrated into the daily curriculum and instructional activities of a classroom” and which they contrast with

“on-demand assessments whether performance-based or not, [which] are not necessarily drawn from the actual repertoire of the classroom” (p. 75).

Portfolios can also be viewed as a type of performance assessment. Portfolio assessment has been a significant aspect of New Zealand primary schools in the last fifteen years and will be discussed in more detail later in this section.

Claims made in favour of Performance Assessment

Linn and Gronlund (2000) identify four advantages of performance assessment. They argue that performance assessment can communicate instructional goals that involve complex performance. This can happen in the midst of normal instructional activity. The authors also see performance assessment as measuring complex learning outcomes that cannot be assessed in other ways. The third advantage is that performance assessment can measure process as well as product. Their final point is the alignment of performance assessment with modern learning theory. “Rather than viewing students as recipients of discrete bits of knowledge, modern learning theory conceives of students as active participants in the construction of meaning” (p. 67).

Eisner (1999) expanded on Linn and Gronlund’s (2000) four points, in particular the measurement of complex skills that children will need in the future.

Performance assessment is a closer measure of our children’s ability to achieve the aspirations we hold for them than are conventional forms of standardised testing. Indeed our educational aspirations have been influenced by the fact that our children will inhabit a world requiring far more complex and subtle forms of thinking than children needed three or four decades ago. For example our children will need to know how to frame problems for themselves, how to formulate plans to address them, how to assess multiple outcomes, how to consider relationships, how to deal with ambiguity. (p.658)

The skills Eisner (1999) talks about are prominent in the revised New Zealand Curriculum (MOE, 2007) and recent books such as *Catching the Knowledge Wave* (Gilbert, 2005).

Support for Claims

Meisels et al. (2001) documented the research literature that supported claims made about performance assessment:

1. Performance assessment improved teaching and learning: Borko, Flory and Gumbo (1993); Darling-Hammond (1994); Falk and Darling-Hammond

- (1993); Gearhart, Herman, Baker, and Whitaker (1993); Khattri, Kane and Reeve (1995); Lin (1993); Meisels (1997); Meisels, Dorfman and Steele (1994); Taylor (1994).
2. Performance assessment resulted in positive changes in instruction: Koretz, Stecher, Klein and McGaffrey (1994); Kentucky Institute for Education Research (1995); Almasi, Afflerbach, Guthrie, and Schafer (1995); Koretz, Mitchell, Barron and Keith (1996).
 3. Performance assessment resulted in better curriculum integration: Khattri et al. (1995).

However, Meisels et al. (2001) also document research that challenged the claims made in favour of performance assessment.

1. Claims reported in favour of performance assessment were challenged: Green (1998).
2. Performance assessment worked against a constructivist approach to teaching: Murphy, Bergamini and Rooney (1997).
3. Examples of inadequate reliability: Linn (1994).
4. Generalisability was limited: Shavelson, Baxter and Pine (1994).
5. Manageability was a problem: Cizek (1991).

NZ policy and other government documents

Mathematics (MiNZC) (MoE, 1992) was one of the first of the seven curriculum documents introduced in New Zealand and, as noted earlier, provided guidance on the increased use of performance assessment. The document also suggested strongly to teachers that more was required than standard paper and pencil tests.

Assessment Policy to Practice (MoE, 1994b) further outlined the Ministry of Education's beliefs about assessment. It included sections on observation and portfolios.

Recent New Zealand Developments

More recently the Ministry of Education has developed a number of assessment activities that support schools developing performance assessment. This includes elements of the ARBs, the New Zealand Curriculum Exemplars (Chamberlain, 2000,

2001), and NEMP Tasks. By definition, all assessment of children's writing is performance assessment. The ARBs, the Exemplars and asTTle Writing all contain guidance on assessing writing performance. Each of these assessment tools also provide a framework for guiding assessment through use of descriptors and guidelines based on levels of the NZCF. As such, they show a relationship between performance assessment and SBA. The ARBs and the Exemplars also provide guidelines for performance assessment in other areas, especially oral language.

The New Zealand Curriculum Exemplars, which began their development in 1999, are not explicitly performance assessment tools, but do guide teachers toward this approach. Chamberlain (2001) describes the Exemplars as: "nationally moderated examples of student work that are annotated to illustrate achievement, and quality in relation to the levels described in the national curriculum statements" (paragraph 1). The Exemplars were developed in all seven curriculum strands from levels 1 to 5. This tool can be seen as being fundamentally based on SBA, in that they attempt to make more explicit the standards and expectations of the levels of the curriculum. However, in their implementation they have strong elements of performance and formative assessment.

In terms of manageability, the question that the Exemplars raise is not whether the tool itself is manageable, but how much effort primary schools should put into assessment of learning in relation to the non-core curriculum.

UK Experience

A key detriment in successful assessment practice is the relationship between the design of the assessment procedure and the purpose of the activity. The experience of attempting large scale performance assessment for accountability purposes in England and Wales illustrates this point. Following the introduction of the Education Reform Act in 1988, the government introduced the Standardised Assessment Tasks (SAT) at ages 7, 11 and 14 years. These assessment activities, which originally had a large element of performance assessment, were designed to be used for diagnostic, formative and accountability purposes. Their implementation and subsequent manageability problems have been well documented (Abbot, Broadfoot, Croll & Pollard, 1994; Dearing, 1994; Gipps, 1995; Torrance, 1993). As Gipps notes, the use of performance assessments for accountability purposes poses major problems.

It is clear the SAT as originally conceived is simply not appropriate for assessing literally hundreds of assessment points; it too time-consuming for testing whole age groups of pupils, particularly at a certain point in time. For what is essentially survey testing, something quicker and more reliable is needed. The SAT model on the other hand is ideal to support individual diagnostic and formative assessment by teachers for their own purposes

The 10-level student grading scale has been replaced by an eight-level scale to try to describe more reliably students' SAT achievement. Finally, in response to teachers' concerns over the time needed to carry out the national assessments, the government is moving away from performance-oriented assessments and replacing these with more conventional, closed-item tests. Consequently, the time required for SAT administration has been diminished significantly. Changes have been made to alleviate the most unrealistic aspects of the Act, and it is likely that additional changes will be made.

(p. 119)

Performance Assessment and the National Educational Monitoring Project

The most significant development in performance assessment in New Zealand began with the establishment of the NEMP in 1995. It differed significantly from the UK approach in that only a light sample (3%) of students were assessed annually and the results were not used to make individual schools accountable. Flockton (1999), one of the directors of NEMP, described the Project as a national *temperature taking* on children's achievement.

NEMP's approach to assessment illustrates a number of challenges faced with a broad approach to assessment for accountability purposes. The assessment is based on a national sample of 3% of year 4 and 8 students drawn from a range of schools. The assessments are administered by experienced teachers trained in the use of these tasks. The teachers work in pairs to administer the tests in all the schools in their designated area. The assessment tasks of the NEMP include a high number of performance assessment tasks and include group as well as individual activities.

Flockton outlines the rationale for the choice of performance of assessment.

Traditional paper and pencil methods have some real limitations for understandings and applications of learning in many areas of the curriculum. The question always needs to be asked: Does the form of assessment help us to tell us about the abilities we are interested in, or does it test something else such as the ability to decode and comprehend a written statement? If we want to know if a student can add two numbers together then a paper and pencil test might be quite appropriate. If we want to know about the strategies they use for solving a real life mathematical problem, then a written form of testing may give an incomplete or misleading account of what they do. (p. 7)

For this reason NEMP places a lot of emphasis on performance assessment. To quote Flockton again:

To overcome the limitations of traditional forms of assessment, national monitoring makes considerable use of performance methods. These methods are intended to engage students in hands-on approaches where the content, processes and materials used are “authentic” in their resemblance of tasks they might experience in real life. Arguably, performance assessment methods have advantages of giving more meaningful and comprehensive information about student ideas and understandings and capabilities in many aspects of curriculum, but like traditional forms of assessment they are not perfect. (pp. 7-8)

The results of the assessments are analysed in various ways and annual reports are given to all schools. By using only a light sample of 3% of students at years 4 and 8 and only assessing three curriculum areas a year, the NEMP project has allowed New Zealand to develop a national testing regime that includes the assessment of performance tasks. The fact that NEMP assessments are conducted by an external agency and selected schools merely supply the students and the space for the activities to take place, means that NEMP assessments are largely unobtrusive for schools and as such have high manageability. Their utility, as far as in individual school are concerned, is more open to question. Schools do not get the results of their students’ assessment and it is left to individual schools to decide whether to use the NEMP reports to influence their programmes. However, it is important to note that the stated purpose of NEMP is concerns a national picture rather than reporting to individual schools as was stated in the NZCF national monitoring is “to provide information on how well overall national standards are being maintained, and where improvements might be needed.” (NZCF, 1993, p26.). Or as NEMP (2002) itself state:

Assessment and reporting procedures are designed to provide a rich picture of what children can do and thus to optimise value to the education community. The result is a detailed national picture of student achievement. It is neither feasible nor appropriate, given the purpose and the approach used, to release information about individual students or schools (p.5)

Portfolios

Portfolios can be seen as one type of performance assessment but are treated separately here because of their widespread use in New Zealand schools and their importance for exploring the manageability and utility of assessment. In examining the history, development and claims made in favour of portfolios, Ward and Murray-Ward (1999) cite Arter and Spandel’s definition as being the most useful.

A purposeful collection of student work that tells the story of the student's efforts, progress and achievement in given areas. This collection must include student participation in the selection of portfolio content; the guidelines for selection; the criteria for judging merit, and evidence of student self-reflection. (p. 193)

Portfolios are seen by advocates of this approach to have a number of strengths. The use of portfolios can be readily integrated with normal classroom instruction.

Portfolios allow students to demonstrate what they can do. They can also encourage students to become reflective learners and develop the skills of self-assessment. Associated with this is the potential for portfolios to help students take responsibility for setting goals and evaluating their progress. Portfolios are also seen as an effective means of communicating with parents on students' achievements (Linn & Gronlund, 2005).

In addition to the strengths listed above, Ward and Murray-Ward (1999) mention that portfolios promote more valid assessment with a better match with the curriculum. They also allow teachers to examine process as well as product. Ward and Murray-Ward also claim that portfolios provide more useful information to teachers.

While using portfolios for assessment and teaching does have recognised strengths, the process is time consuming for both teachers and students. Portfolios also require very clear guidelines to be used effectively. Both the writers quoted above (Linn & Gronlund, 2005; Ward & Murray-Ward, 1999) make a distinction between portfolios and student folders. For Ward and Murray-Ward, a student folder is distinguished from a portfolio in a number of key ways. In the case of folders:

- Teachers control the content and the selection process
- Products reveal completion of a task not process
- There is no evidence of self-reflection by the student

According to Linn and Gronland (2005), it is very easy for teachers and schools to underestimate what is required in a portfolio.

Perhaps the biggest obstacle to realising the potential value of effective portfolio use is not a weakness, per se, but it is the naïve perception that portfolios can be easily created. Since teachers and students are accustomed to keeping work in folders and files, it seems but a simple step to call the folder a portfolio. Unorganised collections of work in folders will not accomplish the goals implicit in the strengths listed in the box. Nonetheless, considerable evidence from surveys, interviews and teacher observations

shows that far too frequently so-called portfolios are indistinguishable from unorganised collections of work in folders with inadequate specifications of purpose, guidelines for construction and evaluation criteria. (p. 292)

As discussed earlier and writing from a New Zealand perspective, Williams (1999), in her study of assessment procedures at a Marlborough school, articulated some of the same concerns as Linn and Gronland (2005) with what were termed *clearfiles*. She said that it was common for no criteria to be set for what was collected, and assessment results were often included but there was no analysis of information. In general, it was common for there to be no school-wide guidelines on what to collect and what to discard from year to year.

Baker (1997), a New Zealand professional development adviser, makes a distinction between student-centred portfolios and standards/level referenced portfolios. She divides the student-centred portfolios into showcase, structured or self-selected. While showcase and self-selected portfolios have a target of teachers, parents and students and aim to celebrate students' achievement, structured portfolios include an element of programme planning and evaluation and have an audience that extends beyond parents, students and teachers to course designers and external auditors. Standards portfolios, according to Baker, are for teachers and include school developed standards and exemplars of students' work. The purpose of standards portfolios are:

- For quality accountability and assurance assessment systems
- For consistent and fair assessment systems
- For improving the quality of school-based assessment
- For promoting a common interpretation of national achievement objectives
- To support the process of self review

(p. 16).

As well as teachers, parents and students, Baker (1997) includes ERO and BOTs as an audience for standards portfolio. In *Assessment Ways Forward*, Baker and Lorrigan (1996) present a model of a standards based portfolio for a primary science curriculum. The implication was that such documents should be developed in all seven curriculum areas; clearly this had serious implication for the manageability of assessment.

One problem creating confusion for assessments in New Zealand primary schools is that two movements occurred at a similar time. The first movement was the NAG

requirement from 1993 for schools to report on individual student's achievement (MoE, 1993c). This led to the development of folders (Linn & Gronlund, 2000) or clearfiles (Williams, 1999). These documents often contained samples of students' work as well as assessment results and were largely kept for compliance purposes to be shown to ERO. It is unfortunate that the compliance requirements coincided with the dissemination of the philosophy underpinning the use of portfolios in assessment and teaching; this led to confusion between the purpose and approach in using portfolios. It is worth noting that teachers in the studies by Hill (2000) and Syme and Timperley (2003) reported that the keeping of students' folders had little value. A more recent development in some schools is the use of electronic or e-portfolios. It is perhaps too early as yet to provide commentary on their use.

Linn and Gronlund (2000) express caution on the use of portfolios in terms of manageability.

Although the potential strength of portfolios makes them attractive both as aids in instruction and as assessment devices, portfolios have weaknesses. They can be quite time consuming to assemble. The investment in students' time in construction of portfolios may be well spent, but teachers need to guard against the tendency for the portfolio demands to foster busy work that contributes neither to greater student learning or better assessment.

Portfolios are time consuming for teachers as well as students. Although students can benefit from the process of constructing a portfolio, they need to have constructive feedback from teachers on the work included in the portfolios and the portfolios as a whole. Considerable thought, preparation and experience is needed to ensure that the benefits of portfolios justify this investment in time by both students and teachers (pp. 291-292).

Summary of Performance Assessment

At this stage in its development, performance assessment is problematic in terms of manageability, particularly when used for large-scale assessment (Gipps 1995). However, the approach does hold promise for improving student engagement and achievement where the goals and procedures are clear and the teachers have had well developed and substantial professional development (Khattri et al., 1998).

3.15 Formative Assessment

Introduction to Formative Assessment

Formative assessment has been the most prominent movement in assessment over the last 10 to 15 years. It is held up as the approach which will make the most significant

difference to children's learning. Formative assessment has been defined by New Zealand researchers Bell and Cowie (1999) as "the process used by teachers and students to recognise and respond to student learning in order to enhance that learning, during the learning" (p. 101). Formative assessment establishes where a student or students are in their learning and shows them the most appropriate next step. Such behaviour has characterised good teaching since instruction first began. Gardner (2006) traces the use of the term *formative assessment* back to Scriven (1967), who distinguished between the timing of assessment in relation to learning: formative assessment is undertaken during the learning, and summative at the end.

Seminal works in the field have been Crooks (1988) and Black and Wiliam (1998). Clark, Timperley and Hattie (2003) summarise Crook's findings to highlight eight ways in which assessment practices impact on student outcomes.

1. Assessment practices affect students in a number of ways. It guides them in making judgments on what is important to learn. It can reduce or enhance their motivation to learn and their perceptions of themselves as learners.
2. Classroom evaluation tends to focus on recall or recognition of isolated pieces of information rather than more global or holistic learning.
3. There has been too much emphasis placed on grading students' work rather than using assessment to assist children to learn. This has reduced intrinsic motivation.
4. Feedback would be more effective if it focused on mastery, was specific and during or soon after the task was completed.
5. Peer assessment and co-operative grouping enhance learning.
6. Evaluation standards should be high but attainable.
7. There is a need for alignment between the assessment tasks and learning goals.
8. Evaluation enhances learning if it includes important skills, knowledge and attitudes. (p. 11)

In 1998, Black and Wiliam carried out a review of literature in the field. They concluded that formative assessment strategies raised standards of achievement. The Black and Wiliam study included an examination of 250 studies. Of these, 30 included quantitative studies based on experimental and control groups which showed

support for formative assessment: “Strengthening the practice of formative assessment produced significant and often substantial learning gains. They ranged over various age groups (from 5-year-olds to university undergraduates), across several school subjects and involved several countries” (p .11).

In the King’s-Medway-Oxfordshire Formative Assessment Project (KMOFAP), Black and Wiliam (2006) successfully implemented the findings of their earlier review in a two year project involving six schools and 48 teachers. The results of the project showed an overall and significant gain in achievement outcomes. The project was based on four activities:

1. Oral feedback in classroom questioning (more recently re-labelled dialogue)
2. Feedback through marking
3. Peer and self assessment
4. The formative use of summative assessment (p. 14)

In one of the few studies of formative assessment that used a quantitative research approach, McDonald and Boud, cited in Broadfoot and Black (2004), demonstrated how a formative assessment approach with a strong emphasis on self and peer assessment showed gains in examination scores

Implementation in NZ

References to formative assessment were contained in a number of Ministry of Education documents:

“The primary purpose of school-based assessment is to improve students’ learning and the quality of learning programmes” (NZCF, MOE1993a, p. 24).

“Teachers should use a range of formal and informal approaches to assessment to take account of student varying needs and styles. These include:

Ongoing, continuous assessment which provides immediate feedback, enhancing the learning as it proceeds.” (EiNZC, MOE, 1994a, p. 21)

The Ministry of Education’s 2001 Assessment Strategy placed a great deal of emphasis on formative assessment. The Ministry’s support of this project is clearly evident on the assessment page on TKI (MoE, 2007b). Support for the approach is also seen in the content of the ATOL and literacy contracts.

The Education Review Office features formative assessment strongly in their *Assessing and Feeding back (Process Indicators)*: “Teachers provide regular, specific and constructive feedback on students’ work that contributes to the next stage of learning” (ERO, 2007).

The implementation of formative assessment in New Zealand classrooms was also enhanced by the publication of a number of less academic type texts aimed at teachers. These books included both British and New Zealand publications. The titles below illustrate the approach and target audiences.

- *Enriching feedback in the primary classroom: Oral and written feedback from teachers and children*, Clarke (2003), Britain
- *Unlocking formative assessment: Practical strategies for enhancing students’ learning in the primary and intermediate classroom, New Zealand edition*, Clarke, Timperley and Hattie (2003)
- *Targeting assessment in the primary classroom: Strategies for planning, assessment, pupil feedback and target setting*, Clarke (2004), Britain
- *Using evidence in teaching practice: Implications for professional learning*, Timperley and Parr (2004), New Zealand.
- *Formative Assessment in Action: Weaving the elements together*, Clarke (2005), Britain.
- *Clarity in the classroom, using formative assessment, building learning focused relationships*, Absolum (2006), New Zealand

Self-Assessment

Black and Wiliam (2006) include self-assessment as a key element in formative assessment. Boud (1986) defines self-assessment as follows:

Self-assessment requires students to think critically about what they are learning, to identify appropriate standards of performance and to apply them to their own work. Self assessment encourages students to look to themselves and to other sources to determine what criteria should be used in judging their work rather than being dependent solely on their teachers or other authorities.

(p. 7)

It could be argued that the basics of formative assessment, the establishment of where a learner is in their progress, and guiding them to the most appropriate next step, has always been a characteristic of good teaching. A similar case could be made for self-assessment. Motivated learners have always made judgements on their own performance in order to make progress in their learning.

Broadfoot (2007) describes self-assessment as much a learning activity as an assessment practice. It is claimed that self-assessment encourages motivation (Broadfoot, 2007; Hill, 1995) and lessens dependence on the teacher. It is also argued that self-assessment should help the students understand the instructional objectives (Linn & Gronlund, 2000). However, Broadfoot makes the point that, for this to happen, there is a need for clarity in setting the standards of performance and she recommends student involvement in this process. This is especially important where self-assessment is used as part of the formal assessment process. A further point that Broadfoot makes is that, where the process does not contribute directly to a formal judgement on performance, students may not take the process seriously.

Hill (1995), working with pre-service teachers, investigated self-assessment in two Waikato schools. While a supporter of the approach, she did express some cautions:

Implementing self-assessment in primary schools is not an easy task. It is one that requires confidence, knowledge and commitment on the part of the teacher, a supportive policy and viewpoint from the school's leadership and enthusiasm from the children. (p. 69)

The Challenges Associated with the Implementation of Formative Assessment

The publication of Black and Wiliam's (1998) *Inside the Black Box* aroused a great deal of interest in formative assessment, first in the UK and then in other countries including New Zealand. Since 1998, Wiliam and Black have been among a number of writers who have attempted to develop a fuller understanding of formative assessment. Gardner's *Assessment and Learning* (2006) discusses recent developments in the implementation of formative assessment. In the chapter *Developing a Theory of Formative Assessment*, Black and Wiliam (2006) use an activity theory model to analyse formative system procedures in their KMOFAB research. They make a number of points that illustrate the complexity of formative assessment when explored over time at a deep level. They note there is great deal of difference between giving feedback in science and mathematics where there is "a body of subject matter that teachers tend to regard as giving the subject more unique and defined aims" and a more open subject like English where there is "no single goal appropriate for all" (p. 85).

Black and Wiliam's (2006) research also makes the point that it takes time to develop the relationships and class culture necessary to successfully implement formative assessment.

More generally, it is important to look broadly at the 'regulation potential' of any given learning activity, noting however that depends on the context, on what students bring, on classroom culture that has been forged 'upstream' (that is, the procedures whereby a student comes to be placed in a context, a group, a situation), and on ways in which students invest themselves in the work. Several of the project teachers have commented that when they now take a class in substitution for an absent teacher, the interactive approaches that they have developed in their own classes cannot be made to work.

(p. 24)

New Zealand researchers Bell and Cowie (1999) comment similarly: "Interactive formative assessment depended on the teachers' skills of interaction with the students and the nature of the relationships they had established with the students" (p. 102). When talking about the changes required in the relationships between teachers and students when formative assessment is implemented, Black and Wiliam (2006) state:

Some teachers have seemed quite comfortable with the transfer of responsibility to the student, and implications for change in the student's role and in the character of the teacher-student relationship are clear. However for some other teachers such changes are threatening rather than exciting. (p. 87)

It is interesting to note that in his recent book *Clarity in the Classroom using formative assessment – building learning focused relationships*, New Zealand educator Michael Absolum (2006) combines the concepts of formative assessment and classroom relationships in line with direction given by Black and Wiliam. Absolum sees the creation between the teacher and the student of what he terms a learning-focused relationship as a key in developing students' learning.

Clarke carried out a large scale research project on the use of formative assessment titled *Gillingham Partnership Formative Assessment Project* (Clarke & McCallum, 2002). While students in the project made progress, they did not differ a great deal from the control group:

Reading: Twelve of the 15 Gillingham schools showed improvement compared with 8 out of the 11 control schools.

Numeracy: Fourteen of the 15 Gillingham schools showed an improvement in their numeracy scores compared to 9 of the 11 control schools. (p. 43)

A number of the teachers, especially those who worked with the younger students reported serious manageability issues in adopting a comprehensive formative assessment project with all children.

Furthermore, Clarke and McCallum (2002) reported it was very challenging to get teachers to fully participate in the programme if they prescribed to a behaviourist, rather than a constructivist view of education. The problems with implementing formative assessment when there is lack of alignment between formative assessment and behaviourist views of education was also noted by Carless (2005).

The Manageability of Formative Assessment

It is perhaps too early to be definite on the manageability of formative assessment, although as noted above, the Clarke and McCallum (2002) study identified some concerns. The widespread use of this approach is still relatively recent in classrooms and over time we may get a fuller understanding of implementation issues such as manageability.

Summary of Formative Assessment

The practices associated with formative assessment do offer evidence that when implemented in a prolonged, structured and supported manner students' achievement improves. Research since Crook's major review in 1988 makes it very clear that the use of formative assessment to improve student achievement is not a simple, quick fix; it requires fundamental change in the way classrooms and schools operate.

3.16 Evidence-Based Teaching

Introduction to Evidence-Based Teaching

Evidence-based teaching has a close relationship to formative assessment. Current understanding of the term *formative assessment*, especially as it used in New Zealand schools, focuses on assessment of individual children, learning intentions, success criteria, feedback and next steps (Clarke, Timperley & Hattie, 2003). Evidence-based teaching is more concerned with the assessment of cohorts of children and has a focus on both groups and individuals within the group. However, it is noted that Reading Recovery (Clay, 1993), which is based on daily Running Records, examination of trends in achievement, and discussion of data with teaching peers and tutors, is an example of individual student, evidence-based teaching.

While it could not be strictly called a New Zealand development, the implementation of this approach has been as much home-grown as introduced from offshore. The fact that, of 38 titles listed in the bibliography of Timperley and Parr's (2004) *Using evidence in teaching practice*, 27 are New Zealand sources, gives a sense of how this concept has developed locally in recent years. The Ministry of Education has allocated considerable resources to develop this approach in schools. In their pamphlet *What Matters Most* (2003b), the MoE state:

By gathering and analysing achievement information, teachers and principals are able to focus on the learning needs of their students, both in the classroom and for the school as a whole. Teaching becomes more focused on making a difference and achievement rises. (p. 1)

From an international perspective, Fullan (1995) argues for “a commitment to data-driven decision-making and using assessment data to support student success as key elements for successful large scale education reform initiatives” (pp. 2-3).

Definitions and related terms: Evidence-Based Teaching

Timperley and Parr (2004) describe evidence-based teaching as, “examining evidence of student achievement for its implications for teaching to raise achievement” (p. 11).

In New Zealand, the approach has included the notion of “mining data” and the need for “professional learning communities” and “learning organisations” to support teaching practice.

Mining data

By *mining the data*, Timperley and Parr (2004) refer to a process that involves a close examination of assessment information to ascertain the impact of current teaching on students' achievement. Such a process can include a close analysis of sub-tests in STAR reading or the use of a wedge graph to track children's reading achievement. It is recommended that these activities are undertaken with groups of teachers rather than individually.

Professional Learning Communities and Learning Conversations

The collaborative discussions among teachers that examine evidence of achievement and provide recommendations for adaptations in teaching approaches to enhance learning, have been termed *learning conversations*. The groups of teachers involved have been termed *professional learning communities*. Timperley and Parr (2004)

make a distinction between professional communities and professional learning communities.

Professional communities share ideas, work together and support one another. When they are working well they enhance the quality of professionals' lives because most of us prefer to work in supportive environments. Unfortunately, there is usually little spill over benefit to student achievement levels. Professional learning communities, on the other hand, also share ideas and support one another, but the interactions among the members of the community are focused on raising student achievement. Every idea is tested for its effectiveness in achieving this end. (p. 115)

Implementation of Evidence-Based Teaching

Evidence-based teaching has been a major feature of a number of recent New Zealand education initiatives. These have included the *Strengthening Education in Mangere, Otara (SEMO)* (Lai & McNaughton, 2003; Mallard, 2003; Phillips, McNaughton & MacDonald, 2001; Timperley, Phillips, Wiseman & Fung, 2003), and post school review developments in Wainuiomata (MoE, 2003). Other recent New Zealand uses of the approach have been reported in Heron et al. (2001) and Symes and Timperley (2003). The MoE belief in the approach is illustrated by the following quote from the Minister's (Mallard) 2003 *Report on the Compulsory School Sector*:

Evidence-Based Enquiry

Recent examination of why some of the SEMCO schools were more successful in raising student achievement than others focused on the characteristics of effective professional learning communities. It appeared that those teachers and leaders who spent their time in professional learning communities examining the impact of their teaching on student achievement and then taking steps to change their pedagogical strategies when students were not responding to their teaching were more successful in promoting achievement. Other schools spent their time discussing implementation issues, but did not test the impact of the implementation on students' achievement. (p. 41)

Challenges and limitations: Implications for the Manageability and Utility of Evidence-Based Teaching

Introducing new approaches to assessment and teaching are often neither simple nor straightforward. While the school leaders in the Symes and Timperley (2003) and Heron et al. (2001) studies were ultimately successful in implementing evidence-based teaching in their schools, they both faced considerable teacher resistance in the initial stages of the project.

The introduction of evidence-based assessment requires some technical knowledge of statistical techniques and a critical knowledge of the tools that are being used to gather evidence. Timperley and Parr (2003) stress the importance of selecting the most appropriate tool. With regard to individually administered tools such as Running Records, they state:

Individually administered tests are very time consuming, and as we suggested earlier, should be completed only when the diagnostic information is to be used to shape and refine programmes for individual students. While they can be used for other purposes, such as developing class or year level profiles, the time involved can rarely be justified if this latter purpose were the primary or only one. (p. 3)

In their September 2003 advice to school, *Planning for Better Student Outcomes*, the Ministry of Education (2003c) present Oruaiti School as a case study of best practice in using Running Records for what they termed analysing student achievement data for improving learning. This advice is somewhat questionable, given the fact Running Records are very time-consuming (Timperley & Parr, 2003; Timperley, 2003) and an inappropriate tool for competent readers beyond year 3 (Timperley, 2003; Baillock, 2004) and are of variable reliability (Fawson, Ludlow, Reutzel, Sudweeks & Smith, 2006).

3.17 International and National Testing

Introduction to International and National Testing

A feature of international education over the last 20 years has been the impact of international studies of education such as the Third International Mathematics and Science Study (TIMSS), the International Association for the Evaluation of Educational Achievement (IEA), Progress in International Reading Literacy Study (PIRLS), and the Programme for International Student Assessment (PISA). During a similar period a number of countries, including New Zealand, have established their own national testing or monitoring systems. In New Zealand this has taken the form of the National Education Monitoring Project (NEMP).

The section begins with a summary of the various international tests that New Zealand has participated in and a brief comment on their impact on education policy and school and classroom practice. This will be followed by an examination of NEMP.

Foster, cited in Doig (2006), claims that “large-scale assessment programmes are designed to assess and monitor the health of an education system and to improve student learning by providing information to stakeholders at different levels of the system” (p. 265). Or as Wagemaker (1993) expresses it:

Comparative perspectives also allow us to examine the impact on educational systems of policies that are applied consistently within nations but may vary across nations. The understanding we obtain from cross national comparisons of such policies as age of school entry, hours and methods of instruction, and teacher training can provide us with new insights into performance of our own educational system in general, and the relationship between student performance and its antecedents and consequences in particular. (p. 1)

Table 3.1: New Zealand International Assessment Results

Date	Survey	Reference	Results
1970	IEA Reading	Wagemaker (1993); Elley (2006)	New Zealand 9 and 14 year olds first in reading achievement among all countries
1980- 1982	Second International Mathematics Survey	Ministry of Science, Research and Technology (1998); Elley (2006)	New Zealand students on average scored around the lower quartile of countries that participated in the survey.
1990	IEA Survey	Smith & Elley(1994); Elley (2006)	Results were not as high as in 1970 but still high: 6 th in reading literacy at year 4, 4 th in year 6. Girls out-performed boys Pakeha/European out-performing other ethnic groups.
1994 1998- 1999	TIMSS TIMSS_R Mathematics and Science	ERO (2000); Elley (2006)	New Zealand performance was below the international mean in Mathematics and slightly above the international mean in Science.
2001	IEA Trends Reading	Elley (2006)	No significant change in New Zealand performance from 1990-2001. Still a large number of high achieving students but a very long tail. New Zealand had one of the highest spreads of achievement.
2000- 2001	PIRLS PISA	Caygill and Chamberlain (2004); Elley (2006).	New Zealand Year 5 students scored 12 th out of the 35 countries. Once again the range of scores was wider than most other countries. PISA 2000 assessed 15 year old students in 32

countries in the areas of reading literacy, mathematical literacy and scientific literacy. New Zealand students had the third highest mean scores for reading literacy and mathematical literacy and sixth in scientific literacy. New Zealand had the highest proportion of students in the highest reading level but also had one of the widest spreads of achievements.

New Zealand's participation in International Assessments

In his 2006 summary of 35 years of international testing, Elley stated how consistent New Zealand results were over time. Our students perform better in reading than they do in mathematics. New Zealand reading results have shown that we have a consistent and large group of children scoring in the highest quartile for literacy. However, at the same time we have had one of the largest spread of scores with a tail of students who are in the lower end of results. The mathematics achievement of New Zealand has been below the international mean on nearly all surveys. That New Zealand performs better in reading than mathematics has a long history. Fifty years ago A.E. Fieldhouse, an education lecturer, later professor, at Victoria University, reported (in Ewing, 1970) that when the results of standardised tests were compared between Australia and New Zealand it was found that New Zealand children were the equivalent of a year ahead of their Australian counterparts in reading but matched all states except Queensland in mathematics. Queensland results were superior to all other Australian states as well as New Zealand (Ewing, 1970).

Government's Responses to International Comparative Studies

In 1997, the Government established the Mathematics and Science Task Force (MoE, 1997c) in response to the TIMSS results. The Taskforce made a number of key recommendations:

- The need to raise teachers' and parents' expectations of success;
- The need to improve the professional skills, knowledge and confidence of teachers;
- The need to provide resources and professional development for teachers to assist and support them in implementing the curriculum;
- The need to lift Maori and Pasifika Islands students' levels of achievement (Higgins, 2002, p. 160).

In 2000, the Ministry of Education introduced the *Count Me In Too* professional development programme, which later grew into the Numeracy Project in response to the low performance of New Zealand students in the 1994 TIMSS study (Higgins, 2002). This was one element of the Literacy and Numeracy Strategy. The Literacy and Numeracy Strategy had three key goals: clarifying expectations, lifting professional capacity, and developing community capacity. Thomas and Ward (2001)

also support the view that government agencies viewed the low performance in mathematics as largely the result of low teacher capacity in the area.

The responsibility of poor international standing of our students has been widely attributed to the lack of quality in our mathematics teaching (Garden, 1997; the New Zealand Ministry of Education, 1997a; the New Zealand Ministry of Education, 1997b; and the Education Review Office, 2000) (p. 1).

Higgins (2002) credits the development of the Numeracy Project to poor results in TIMSS. She also cites problems with the implementation of MiNZC, as noted by ERO (1994a) and Holton et al. (1996), and poor student achievement in mathematics as identified by Flockton and Crooks (1997). The TIMSS results were also one of a number of factors that led to the changes to the NAGs in 2000, giving more emphasis to numeracy and literacy.

Other less direct examples of international and national assessment results influencing government policy can be seen in relation to the initiatives that focus on New Zealand's large tail of underachievement, such as the Strengthening Education in Mangere and Otara Project (SEMO) (Phillips, McNaughton, & McDonald, 2001), ICAN Porirua, and Taita Central School Literacy project (Mallard 2000; Mallard, 2003). Knight (2000) argues that there is a connection between New Zealand's poor performance on the TIMSS study and the national testing debate of the late nineties (MoE, 1998), as well as the requirements for the annual reporting of achievement that were a requirement of the Education Standards Act (MoE, 2001).

International Assessment and Manageability

International testing makes very few demands on schools directly in terms of the manageability of assessment. However, it could be argued that assessment demands associated with the Numeracy Project can be traced to New Zealand's poor results in the TIMSS assessment in the mid-nineties (Higgins, 2002). As stated in the previous paragraph, Knight (2000) presents the view that these poor results also influenced the introduction of the annual assessment requirements that were made mandatory with the Education Amendment Bill No 2 (2000).

National Assessment Regimes: The National Educational Monitoring Project

NEMP was introduced to provide "a broad picture on achievement for accountability purposes" and "it also makes available detailed information which policy makers, curriculum planners and educators can use to inform and review educational policy,

practice, outcomes and resourcing” (Flockton, 1999, pp. 3-4). Further details on NEMP can be found in Chapter 4.

The NEMP project illustrates some of the dilemmas associated with using performance assessment for accountability purposes. The English SAT system includes large elements of performance assessments carried out in all schools. The tests are administered by the teachers themselves under the guidance of standards and assessment leaders. The results are used in a name and shame league table manner (Thrupp, 2004). As such the system has been criticised for encouraging teaching to the test, creating low teacher morale, being time consuming, and lacking reliability (Gipps, 1994; Elley 2004). NEMP avoids a number of these issues by only assessing 3% of the national cohort of year 4 and 8 students. As such the system is less expensive and time consuming. The choice to use sampling rather than test all students allows the assessment to include a high number of performance tasks and the use of some expensive equipment like videos, DVDs and computers. The English system does, however, provide assessment information to individual schools on the progress and needs of the cohort of children in their school. Such information is not provided in the New Zealand systems where schools only receive information on the national results. Therefore, while the NEMP results may provide valuable information on national trends and the performance of special populations for policy level decisions, the utility of this information for individual schools is open to question. However, in terms of manageability, NEMP places very few demands on schools and teachers. Those schools chosen to participate in the project merely supply the children and a space for the assessment to take place.

3.18 Summary Chapter Three

Chapter Three addresses many of the issues relevant to research question one:

What were the major influences and challenges to assessment in New Zealand primary schools since 1990?

Table 3.2 summarises some landmarks in assessment over the period under investigation.

Table 3.2: Assessment in New Zealand primary schools 1989-2006: some relevant landmarks

Year	Initiative/Publication
1989	<ul style="list-style-type: none"> • Tomorrows' Schools' Legislation
1990	<ul style="list-style-type: none"> • ERO carry out their first review
1992	<ul style="list-style-type: none"> • Mathematics in the New Zealand Curriculum introduced
1993	<ul style="list-style-type: none"> • The New Zealand Curriculum Framework introduced • The National Education Guidelines introduced • Science in the New Zealand Curriculum introduced • First Assessment for better learning (ABeL) professional development contracts begin. The name later changes to Assessment to Learn (AtoL)
1994	<ul style="list-style-type: none"> • English in the New Zealand Curriculum introduced • Assessment Policy to Practice published • TIMSS results published <i>New Zealand performance was below the international mean in Mathematics and slightly above the international mean in Science. This led to the introduction of the Numeracy Project</i>
1995	<ul style="list-style-type: none"> • Technology in the New Zealand Curriculum • Assessing students achievement – <i>ERO provides the first major critique of assessment of the NZ Curriculum</i> • National Education Monitoring Project begins • Social Studies in the New Zealand Curriculum introduced
1996	<ul style="list-style-type: none"> • First assessment resource bank items made available to schools
1997	<ul style="list-style-type: none"> • School Entry Assessment introduced
1998	<ul style="list-style-type: none"> • Assessing Children's curriculum achievement – <i>ERO argues for national testing</i> • Assessment for Success in Primary Schools – <i>National Governments discussion document on national testing</i> • Black and Wiliam publish <i>Inside the Black Box</i> – <i>leading to a subsequent large emphasis on formative assessment.</i>
1999	<ul style="list-style-type: none"> • Health and Physical Education in New Zealand Curriculum introduced • Report of the Literacy Task Force – <i>too much assessment is affecting teaching</i> • Revision of the NAGs – <i>Response to literacy task force – concentrate on literacy and numeracy in years one to four</i> • Information for Better Learning - <i>Plans for national testing – pilot of externally referenced tests in literacy and numeracy- NEMP and ARBs to be expanded</i> • Change of Government – National tests to be developed but their use to be optional
2000	<ul style="list-style-type: none"> • The Arts in the New Zealand Curriculum introduced • Report to the Ministry of Education: A review of the roles and responsibilities of the Education Review Office –<i>ERO required to advise and assist as well as review</i> • Count me in too – <i>later the Numeracy Project begins to roll out in New Zealand schools</i> • The National Assessment Strategy released
2001	<ul style="list-style-type: none"> • Education Standards Act – <i>target setting, planning and reporting</i> • Supplementary Test of Reading Achievement (STAR) released
2003	<ul style="list-style-type: none"> • asTTle released • The Exemplars released
2006	<ul style="list-style-type: none"> • Draft revised curriculum released • Revised PATs released

The last 15 years of primary education has been a challenging journey for New Zealand schools. The first half of this period was extremely difficult as schools attempted to make sense of, and develop, systems for the rapid introduction of a curriculum that was challenging in structure and hard to manage. This struggle took place without the development of appropriate tools to assess the new curriculum, with very little advice and guidance, and the challenging first experiences of highly public ERO reports. Since the millennium there has been an improvement in the support offered to schools and the development of more appropriate assessment tools. Schools are being encouraged to co-operate with their neighbours rather than compete. There remain many challenging aspects to assessment although it appears that the area has improved for schools in the least seven years. This includes the development of new assessment tools and more guidance given to schools.

During the period 1985-2005, approaches to assessment in New Zealand schools were heavily influenced by movements in assessment and education in general that were happening throughout the western world. A major influence was the concept of outcomes-based education that dominated the reforms relating to the governance and management of schools and associated changes in curriculum and assessment in the late eighties and early nineties. Other major influences were performance assessment and the growth of international testing. During the last ten years, the culture, language and practice of formative assessment has dominated schools. More recently, the associated approach of evidence-based teaching has been implemented in a number of New Zealand schools.

Research Question One will be addressed more comprehensively in Chapter Seven.

CHAPTER FOUR: ASSESSMENT TOOLS AND PROCEDURES USED IN NEW ZEALAND PRIMARY SCHOOLS

4.1 Introduction to New Zealand Assessment Tools

This chapter reviews assessment tools currently in use in New Zealand schools with a particular focus on the most pervasive tools such as Running Records and those most recently developed: STAR (Elley, 2001); asTTle; The Numeracy Project; the Exemplars; ARB; School Entry Assessment (MoE, 1997a) and the revised PAT Tests (Darr, Neill & Stephanou, 2006; Darr et al, 2008). While an examination of individual tools and procedures illuminates some general issues in assessment, examining assessment across a particular curriculum area gives a more comprehensive understanding of both manageability and the influence of policy on practice. For this reason, the chapter will conclude with a discussion of assessment in mathematics and the visual arts, so as to illustrate the manageability issues faced by a school in a particular curriculum area.

The development of the tools will be considered in terms of their intended purpose, structure and implementation. Where the test or procedure has been critically evaluated and subject to research this will be discussed. Finally, the assessment tool will be examined in terms of manageability and utility.

4.2 Progressive Achievement Tests (PATs)

The first PATs in Reading Comprehension and Reading Vocabulary (Reid & Elley, 1991) were developed by NZCER in 1969. This was followed by Listening Comprehension (1971) (Reid, Johnson & Elley, 1994), Mathematics (Reid & Hughes, 1974), and Study Skills (Reid, Croft & Jackson, 1978). The Essential Skills (Croft, Dunn & Brown, 2000) replaced the Study Skills and although they were not strictly PATs, they had a similar format and were also developed by NZCER. This organisation also developed the controversial and now discontinued Test of Scholastic Abilities (TOSCA) (1981).

The PATs are norm-referenced, multi-choice tests, although the Reading Comprehension test also contained some criterion-referenced information. They have been widely used in New Zealand primary schools. Dunn, Strafford and Marston (2003) reported that over 90% of schools use the Reading tests and 80% the Mathematics test. ERO (1999) reported 70% of the 118 schools in their sample used PAT Reading and 40% used the Mathematics test. However, in the Dunn et al. study teachers consistently rated the PAT Listening, Reading and Mathematics as the least useful of the assessment tools used. There have been a number of major revisions of the tests. The Mathematics PAT Test was revised in 1994; and the Reading Tests in 2008. The most recent revisions have been Mathematics (Darr, Neill & Stephanou, 2006) and Reading (Darr, McDowall, Ferral, Twist & Watson, 2008).

The Test Manual of the PAT Reading Comprehension and Vocabulary Tests (Reid & Elley, 1991) lists the purposes that the PAT Tests might be used for:

1. To group children with similar needs within a class;
2. To set realistic goals and set materials for students;
3. To identify readers requiring special diagnostic and remedial treatment;
4. To select able readers to be considered for special enrichment programmes;
5. To identify students who are not working up to expectation or who are performing erratically;
6. To provide a frame of reference within which teachers can evaluate the class work of their students;
7. To locate areas of weakness and strengths within a class;
8. To help classify school entrants from outside the New Zealand school system;
9. Broad diagnostic uses: differences in performance between factual and inferential items, rates of reading.

(p. 3)

While there has been a number of articles written about PAT tests the majority have focused on giving teachers greater understanding of the tests and how they can be used in the class and school. Publications with this intention include: Study Skills (Croft, 1984; Reid, 1977) and Information Skills (Brown, 2000; Dunn, 2001). Dunn

(2001) looked at performance in terms of gender and decile rating; Joyce and Darr (2004) provided guidance on using the Teachers Manual.

An interesting investigation was carried out in the use of PAT data to improve student performance at a low decile, South Auckland school (Heron et al., 2001). In this study, PAT assessment data was analysed and used to plan programmes and target the raising of achievement levels. While ultimately successful in getting teaching staff to analyse and use PAT data to enhance student achievement, the campus director at Southern Cross Middle School initially faced challenges from a number of teachers. Staff stated that the tests were “culturally insensitive, inappropriate for decile one schools, and too hard for these children” (p. 35). Such views have been persistent in New Zealand schools for a number of years, particularly in schools with high numbers of Maori and Pasifika students.

The PAT tests have all been reviewed in the *Mental Measurement Year Book: Reading* (Rudman, 1995); *Mathematics* (Brody & Lane, 1995); and the *Essential Skills* (Johnson, 2004). All three tests received very positive reviews with some suggestions to enhance the reliability information in the *Essential Skills*.

Technical Information PAT Tests

Reading Tests

The PAT Reading Tests Manuals (Reid & Elley, 1991) provide extensive technical information on the construction of the tests, norming procedures, reliability and validity. Both tests report at least satisfactory levels of reliability using a variety of statistical techniques; the lowest value obtained was .85. The development of the PAT Reading tests and their later revision involved wide use of literacy advisers and teachers who contributed to the content validity of the tests. The tests have also been shown to have a high correlation with other reading tests such as Tests of Reading Comprehension (TORCH), Iowa Tests of Basic Skills and Burt Word Reading with most tests showing a correlation of between .7 and .89.

Mathematics Test

The 1993 revision of the original PAT Mathematics Test (Reid, 1993) was based on a norm sample of 1000 children at each year level. Reliability measures undertaken for equivalent forms, split-half and the application of the Kuder-Richardson Formula give estimates of above .84 with most measures being above .90.

The validity of the revision of the PAT Mathematics test was enhanced through the involvement of academics, mathematics advisers and teachers. The Test had a correlation of above .75 with equivalent international measures at standard four (year 6) and above, and .67 for year 5.

The most recent revisions of the PATs (Darr, Neill & Stephanou, 2006; Darr, McDowall, Ferral, Twist & Watson, 2008) contain two significant developments. Both these tests are capable of being marked electronically. For a charge of \$ 3.00 per student NZCER will mark the test and produce reports for the school. This development makes the assessment more manageable in terms of time. However there is, as noted, a cost involved. The second major development is the introduction of a scaled score to both tests. This development allows a student's progress to be measured; a feature that earlier forms of both tests lacked.

4.3 Running Records

It could be argued that Running Records are the most widespread assessment practice in New Zealand schools, in terms of the almost universal use of the tool and the time devoted to their administration. Wylie (1999) reported that the approach was used by 96% of teachers at years 4-6 and 83% at years 7-8. The approach is endorsed by the Ministry of Education (2000c) and it is a Ministry expectation that teachers should be familiar with the approach.

Running records were introduced to New Zealand from two sources in the 1970s and had become firmly established by the 1980s. The process can be traced back to Clay's work in the 1970s and a very similar technique, miscue analysis introduced from America by the Goodmans. Smith and Elley (1994) describe the relationship between Running Records and miscue analysis:

Dr Marie Clay designed the procedures for conducting Running Records during the 1970s and her manuals are recommended for those who wish to know more (Clay, 1985). While she was constructing and refining the technique, the Goodmans were simultaneously developing, in the United States, the miscue analysis, which essentially achieves the same ends, using different conventions (Goodman & Burke, 1972). For miscue analysis, the teacher records miscues (or errors) on a duplicate copy of the child's text. No ticks are required for correct responses. Miscue analysis also will typically require pupils to retell what they have read or respond to questions. The subsequent analysis of the two techniques is similar. (p.102)

Briefly stated, Running Records involve teachers monitoring, by a series of ticks and other symbols, a child reading a graded text aloud. The teacher keeps a record of errors and self-corrections. The results are then analysed in a number of ways; the most notable being the establishment of an accuracy rate with above 90% of words correct being described by Clay as at instructional level, and below 90% too hard for the child. It is common for a teacher to get the child to read a number of texts before an instructional level is found. It is also common for the running record to be analysed in terms of the self-correction rate which is the number of errors compared to the number of self-corrections. In junior classes particularly, the child's errors and self-corrections are analysed in terms of the strategies that the child is using and neglecting. For example, does the child use meaning, grapho-phonetic or structure clues when they are attempting to decode text? Although not part of Clay's procedures, it is now also reasonably common for some form of assessment of comprehension to be used in conjunction with Running Records either by having the child retell the passage or having them answer a set of comprehension questions. This is more common with Running Records taken with older children. It is estimated that the time required to administer a running record ranges between 10 to 30 minutes (Timperley, 2003).

The Ministry of Education, in the resource they distributed to all schools (MoE, 2000c), listed the following purposes for carrying out a running record:

- to plan teaching based on what a particular learner attends to when they read and what is appropriate for the next step;
- to identify and observe the particular difficulties of specific students;
- to monitor the progress of individual students as they learn to use a range of information and strategies to make meaning from text;
- to track the progress of individual students as they learn to read texts of increasing difficulty;
- to decide how to group students for instructional reading

(p. 7)

While the administration of Running Records is close to universal practice in New Zealand primary schools, this is not the case in North America. However, it is interesting to note that Pressley, Wharton-McDonald, Allington, Black, Morrow and Tracey (2001) found that taking Running Records at a class level was one of the

hallmarks of effective teachers of literacy. Ross's (2004) research further supports this view.

Recently the widespread use of Running Records has been challenged by two New Zealand educationists (Blaklock, 2004; Timperley, 2003). Timperley's study involved eight Auckland primary schools that required their teachers to carry out at least one running record a term. She concluded that 20% of instructional reading time was taken up with the administration of Running Records, with the average time for the administration of each assessment being 17 minutes. Teachers in the study reported that the most common reason for undertaking a running record was to establish an appropriate text level and group children. In all of the schools in the study, data from Running Records was given to the school's management team where it was used for, among other things, reporting school-wide achievement to the BOT. Timperley questioned the very limited feedback given to teachers on the results of the Running Records by senior management and virtually no feedback from teachers to individual children. Both Timperley and Blaklock question the appropriateness of administering Running Records to children beyond year three. In the words of Timperley:

Running Records are a potentially powerful diagnostic tool for beginning and struggling readers, but it does not follow that learning for all readers will be enhanced through their use. Administering an individualized oral reading assessment for readers "who are still learning about what print is for and how to gain meaning from it" (Ministry of Education, 2000, p. 4) to fluent silent readers is difficult to justify in terms of the time involved and the learning benefits. (p. 73)

In questioning the appropriateness of using Running Records for assessing older fluent readers, Blaklock (2000) finds the comments of Clay (2000) and the MoE (2000a) unclear:

Clay notes that "if Running Records are to be used with older readers there should be a special reason for taking them" (p. 25). However Clay does not elaborate on what that special reason should be.

The Ministry of Education (2000) states that Running Records "can be used at almost any stage of reading development" but are "most useful to teachers of readers who are not yet fluent. This includes most students in years 1 and 2 as well as many older readers" (p. 6). However, the Ministry guidelines do not specify which older readers would be considered "not yet fluent". This group could be interpreted as including all students who are reading below average for their age level. The lack of clear guidance means that schools and

teachers may differ in decisions about which students should be assessed with Running Records. (p. 242)

A further point that requires greater clarification is whether the text used in the running record should be what is termed *seen* or *unseen*; a seen text is one that has been read before by the child and an unseen text is new to the student. Blaiklock (2000) found Clay and the Ministry unclear on this point.

Blaiklock (2000) argues that the use of Running Records for summative assessment purposes is also questionable because of the dubious levelling procedures used to obtain reading ages in both commercially produced and school developed Running Records resources. An additional issue in using Running Records for summative assessment is the lack of reliability in measuring the students' scores in this assessment. Fawson, Ludlow, Reutzel, Sudweeks and Smith (2006) demonstrated that it required the reading of three texts at the same difficulty level to obtain an accurate running record.

ERO (1999) raised issues about the need for greater consistency in the administration of Running Records if this procedure was to be used reliably for summative assessment:

Many schools used information from Running Records to track achievement trends and to monitor students' progress. Teachers in different schools, and often even different teachers in the same school, followed different practices. As well they had only rough information about the difficulty level of the prose they used for testing. Information that is more useful for these summative purposes could be readily obtained if teachers had access to better assessment tools. (p. 37)

While the MoE's (2000a) publication *Using Running Records* guides schools towards the use of Running Records for diagnostic and formative assessment, their 2003 *Planning for Better Student Outcomes* provides an example of two schools using Running Records for summative purposes. By implication, the MoE did not see the lack of reliability in running record assessment as problematic when "the procedure is used to report the progress of individual students and aggregate student progress to the school's wider community" (p. 3).

As stated above, it is also common for Running Records to be accompanied by some form of evaluation in terms of reading comprehension either by asking the child to retell the story or, in the case of a number of commercially produced inventories such

as the commonly used *Probe* (Parkin, Parkin, & Pool, 2002) and *PM Benchmark* (Smith & Nelley, 2002), having the child respond to a set of oral questions. This process does raise questions of the validity of this type of assessment. The problem with such an approach is that the questions have not been normed. Furthermore, there is no evidence of any analysis of the difficulty of the questions and no scientific basis is provided to support the claims in the case of *Probe* that the questions categorised as being inferential, factual and vocabulary do actually represent those domains.

In summary, Running Records are a powerful diagnostic and formative assessment tool for children learning to read. In terms of manageability, their use beyond that stage and purpose is questionable and has, as Timperley (2003) concluded, encroached on a great deal of teaching time. However, given the reporting requirements that were imposed on schools after the education reforms, with little support or guidance, it is understandable that the use of an assessment tool that was familiar to teachers was expanded and used frequently for summative reporting. This is despite the fact that there was a strong argument that the use of Running Records was inappropriate for older children and was administered inconsistently by different teachers. Furthermore, when such a procedure becomes embedded in school culture and practice there is a risk that changes to the practice or even discarding the particular tool altogether will be a challenge for school leaders.

4.4 National Education Monitoring Project (NEMP)

The National Education Monitoring Project, which has been discussed earlier, began in 1995. The assessment is based on a national sample of 3% of year 4 and year 8 students drawn from a range of schools.

The results of the assessments are analysed in various ways and annual reports are given to all schools. As they are a national assessment procedure, NEMP tools were not originally designed to be used by individual schools. However, since 2001 a number of the assessment tasks have been made available to schools. Gilmore (2003) recorded that between 1995 and 2003 approximately 600 “released” tasks have been made available. Dunn et al. (2003) recorded these tasks being used by 12% of teachers at year 5 and 19% at year 7. However, as NEMP is based on year 4 and year 8 students, the use may be higher at those levels. The low uptake of use of NEMP tasks was also evident in a study by Gilmore (2002). She concluded that teachers do

not often make use of the NEMP resources because of the lack of connection between the tasks and the curriculum. In Gilmore's 2002 study, teachers who were task administrators in the 2000 NEMP assessments undertook a curriculum mapping exercise which located 400 NEMP tasks to strands and levels of the curriculum. On a different tack, Gilmore (1999) also recorded that the involvement of teachers in the NEMP process as task administrators greatly increased their confidence in assessment and this was taken back to their schools.

The NEMP website (2007) contains comments by two panels of eminent academics in the field of assessment who have reviewed the Project. Both panels recorded positive conclusions about NEMP:

The National Educational Monitoring Project is well conceived and admirably implemented. Decisions about design, task development, scoring, and reporting have been made thoughtfully. The work is of exceptionally high quality and displays considerable originality. We believe the Project has considerable potential for advancing the understanding of, and public debate about, the educational achievement of New Zealand students. It may also serve as a model for national and/or state monitoring in other countries (Professors Paul Black, Michael Kane, & Robert Linn, 1996).

We want to acknowledge publicly that the overall design of NEMP is very well thought through. The vast majority of tasks are well designed, engaging to students and consistent with good assessment principles in making clear to students what is expected of them (Professors Elliot Eisner, Caroline Gipps and Wynne Harlen, 1998).

While NEMP involved the implementation of the new curriculum the results of assessment tasks were reported as the percentage correct or achieved for each assessment task per year level; and there was no engagement with the levels of the curriculum or specific achievement objectives. Flockton (personal communication, 28 June, 2007) reported that there was some pressure from the Ministry in the initial stages of the design of the project to base items and results on the curriculum levels, but this was resisted by the developers because of their lack of confidence in the level structure.

4.5 The Assessment Resource Banks (ARB)

In April 1996, the first ARBs became available to schools. They were the first assessment tools based on the new Curriculum Statements in Mathematics (1992), Science (1993) and English (1994). For New Zealand they were revolutionary in that they were computer based and only available on-line. The ARBs were developed by

NZCER under contract from the MoE. By 2004 there were 2452 ARB items. The ARBs include a number of items that have formed part of NEMP assessments, Programme for International Student Assessment (PISA), and PATS. Each individual ARB item is trialled on 200 children from a range of eight schools. The developers of the Assessment Resource Banks list seven uses of the ARBs:

Formative Assessment

Diagnostic Assessment

Summative Assessment

Pre and Post Tests

Confirming and reporting levels of achievement

Monitoring and evaluating school-wide performance over time

Exemplars for teacher-made assessment

(Croft, 1999, pp. 32-33)

Dunn et al. (2003) reported 21% of teachers at year 5 and 24% of teachers at year 7 were using the ARBs for assessment in English. This compares with 88% and 90% using PAT reading tests. In mathematics, reported use was 34% (year 5) and 39% (year 7). This compares with 80% (year 5) and 84% (year 7) for Progressive Achievement Tests (PAT) Mathematics. A limitation with this research was that it did not measure how frequently the teachers were using the ARBs. Gilmore and Hattie (2000) reported that in their investigation 43% of schools were registered with NZCER to use the ARBs. In those schools 10% of teachers were high or very high users and 55% were low or very low users. When these data were integrated into schools as a whole it meant that in 2000, 75-85% of New Zealand teachers were low or non-users of the resource. Reported numbers of teachers visiting the ARB site in 2002 showed the hits varied from 3000 per week in the school holidays to 13,000 during term time (Christina Smitts, NZCER, personal communication, 20 December, 2004).

Dunn et al. (2003) also asked teachers and school management to rate the ARBs on a four point scale for how useful they were (1 least useful to 4 most useful). In English at year 5 the ARBs were rated by teachers at 3.3 as useful for providing information. This was below reading prose inventories (3.7) and conferencing and interviews (3.7) but above such common assessment tools as PAT Reading (2.8), BURT (3.0) and the

new reading test STAR (3.0). The results at year 7 were similar. The results for mathematics and English were similar although the difference in both teacher and management ratings of the different assessment tools was not as marked. However, school management (principals, deputy and assistant principals) did not rate the ARBs as highly as classroom teachers. At year 5 they were given the lowest rating (2.3) out of 14 assessment tools; this was below STAR (3.0), PAT Reading (2.9), and BURT (2.8). However, for school management, the major focus is on summative assessment for accountability purposes such as reporting to Boards of Trustees, the Ministry of Education, and ERO. The differences in rating tend to suggest that the ARBs are rated higher by teachers whose focus is more on formative assessment than by management with a focus on summative assessment.

The ARBs give extensive coverage to some areas of the curriculum that are not covered well by other assessment resources. Apart from the NEMP and the Exemplars, the ARBs are the only New Zealand assessment tool that covers the science curriculum. Furthermore, the ARBs give more extensive coverage of the science curriculum than do NEMP and the Exemplars. In the English Curriculum, the ARBs have 160 items in the area of visual language and 82 in oral language.

The ARBs web site has three case study examples of school-wide use of the bank. Upper Hutt Primary School is used as an example of school-wide use of the ARBs in English and Mathematics. However, since this case study was written the school has now abandoned these procedures because they were too time consuming and they believed asTTle delivered better quality information in a shorter amount of time (Peter Durant, Principal Upper Primary School, personal communication, 9 February, 2005).

Most writing on the ARBs has been produced by the NZCER itself with a focus on making the best use of the resource such as the case studies mentioned above. Croft (2001) describes development of the resource. There have also been articles that provide more in-depth analysis of students' responses (Marston & Kenneally, 2001; Neill, 2000).

The ARBs were subject to three evaluations commissioned by the MoE in 2000 (Gilmore & Hattie, 2000; Ham, Schwier, & Davies; 2000; Mendelovitis, Fartoka & Lindsey, 2000). The Mendelovitis et al. study focused on the assessment methodology

used in the ARBs and recommended the continued development of more assessment tasks which required responses other than pencil and paper. They also recommended that items needed more trialling with students from different year levels and that the achievement results from these trials be included in information regarding the difficulty of each item. A further recommendation was for more items to be included from the oral strand of the English Curriculum. Overall Mendelovitis et al. found the ARBs to be based on a well-developed methodology and an exciting development. Ham et al. (2000) evaluated the tool as an electronic resource. They also saw many positive aspects and recommended a number of technical and layout changes.

The Gilmore and Hattie evaluation (2000) concluded that there was a need to increase the use made of the resource by teachers. They noted that high users of ARBs were attracted because of the perceived usefulness for classroom practice rather than ease of access. In a subsequent article Gilmore and Hattie (2001) made some relevant points regarding the manageability and utility of assessment using the ARBs. They stated that time was perceived as a barrier by low or non-users of the ARBs but not by those teachers who used the resource frequently.

4.6 School Entry Assessment

The School Entry Assessment Kit (SEA) (MoE, 1997a) was introduced to New Zealand schools in 1997 and designed to be administered to children in their first six weeks of school. SEA consists of individual assessment in numeracy, oral language and emerging literacy. Schools were expected to send the results of their SEAs to the Ministry of Education. The estimated time for carrying out the assessment is in the vicinity of 45 minutes per child.

The MoE stated the purpose of the Kit was to:

Identify some of the key knowledge and skills of new entrants which will be used as a basis for the planning of programmes;

Provide information for school management about their new entrant cohort to assist in planning and allocation of resources;

Monitor student progress and analyse barriers to learning;

Provide aggregated information to the MoE to inform national policy development. (p. 4)

Goldring (1999) investigated the use of the kit in Christchurch primary schools. In her interviews with teachers using the kit, she identified major manageability issues. “Comments about the time needed for administering the assessment appeared repeatedly. Most teachers found it impractical and impossible to administer it in the classroom without the child and the rest of the class being compromised in some way” (p. 41).

In a similar study in ten Auckland Schools (Dixon & Williams, 2000), teachers reported a number of issues with SEA concerning the procedure’s manageability:

Despite the fact that SEA was designed to be carried out in classrooms, all of the teachers interviewed, without exception, agreed it was impossible to administer SEA effectively within the classroom situation. This was partly because of the time involved in administering the tasks and because of the difficulty in working uninterrupted with one new entrant child for a prolonged period of time while supervising all other children in the class. (p. 33)

For this and other reasons, only seven out of ten schools in the Dixon and Williams study administered all of the SEA tasks.

In 2001 the Ministry of Education commissioned The Australian Council of Education Research to research schools’ use of SEA (Dewar & Telford, 2003). This research concluded that SEA was used by 59% of schools in their sample. Only a third of schools were sending data to the Ministry of Education and only 46% of schools reported they always used the oral language element of the assessment. The oral language assessment *Tell Me* was considered to be time-consuming and difficult to administer. The majority of teachers felt that SEA needed to be up-dated or modified. On a more positive note the majority of teachers agreed with the statements below:

- SEA helps teachers to set appropriate expectations for new entrants
- SEA information is helpful in early discussions and interviews with parents and caregivers about how well a child is getting on at school and what support they may need
- SEA provides the teacher with a representative snapshot of a child’s time at a specific time. (p. 46)

The implementation of SEA illustrates two major themes of this thesis. Firstly, manageability is a major issue in the successful implementation of any assessment procedure. Failure to take account of the time required for carrying out an assessment

activity and the reality of classroom life will lead to less than successful implementation of an assessment activity.

The second point is more subtle. While not mandatory, the MoE requested that schools send the results of SEA to them. This gave SEA an aspect of compliance that other assessment procedures, other than six year nets (schools are required to send annual six-year net results to the Ministry to access funding for Reading Recovery), do not have. Yet despite this compliance aspect only 59% of schools were carrying out the assessment and only a third were sending in their results (Dewar & Telford, 2003). It could be suggested that this response to the Ministry's expectations was the result of schools being given little direction in terms of assessment since the reforms. It could possibly be interpreted that if schools were expected to be self-managing in assessment, this self-management could be extended to ignoring the Ministry's requests for information.

4.7 STAR

The STAR test was developed by Warwick Elley in response to the 1999 MoE's *Literacy Task Force Report*. The Task Force attempted to define what it meant to be a reader. STAR is a group test of reading and includes sections on Word Recognition, Sentence Comprehension, Paragraph Comprehension and Vocabulary Range in years 3 to 6. The years 7 and 8 test also include assessments on children's understanding of the language of advertising and different genres or styles of writing. Elley described the test as being supplementary to the teacher's own observations. He outlined six specific purposes for the test:

1. Help teachers to identify children who need extra help.
2. Help teachers to group children with similar reading needs and abilities.
3. Help teachers to assess the ability of pupils new to the school.
4. Help teachers to diagnose difficulties which are unique to individual pupils, or common to many pupils.
5. Help schools to evaluate a new programme or policy by testing children before or afterwards
6. Provide validation for teachers' judgments about pupils' abilities relative to those in other parts of the country. (p. 8)

The results of STAR are reported in raw scores or stanines. The test also reports typical or expected scores and critical scores. These expected or typical scores show the range where the middle 50% of the students fall. Below the typical score is what is defined as the critical score which indicates in Elley's words *action is needed to be taken*. The typical and critical scores are included in each of the sub-tests as well as the test as a whole. In this way the test is able to profile different strengths and weaknesses of individuals and groups of students. Related to these profiles, such as *high word recognition and low comprehension*, is brief guidance on how to address these issues:

In STAR year 4-6 about 12% of pupils score 90% or more (9 or 10) on Word Recognition, but less than 50% on the other sub-tests of meaning. By contrast less than 5% of year 7-9 pupils show this pattern. These pupils have mastered the elements of decoding, but need more practice at reading interesting text, at their level. They probably also need plenty of encouragement, as they usually lack strong motivation to read by themselves. (p. 24)

However, Elley cautions against putting too much emphasis in the typical sub-test patterns.

Each sub-test measures a different aspect of pupils' reading ability. Therefore, it should be possible, in theory, to identify pupils who have different strengths and weaknesses in reading, and to diagnose and remedy their difficulties. Unfortunately, reading ability is not like a medical condition, which can be neatly classified and remedied with a prescription. All reading skills which are assessed by a group test such as STAR are positively correlated and tend to develop together as pupils practice their reading in different contexts. Pupils who are weak at decoding (Sub-Test 1) tend to be weak in all aspects of reading. Those who can comprehend well what they read in short contexts (Sub-Test 2) are also good at comprehending in longer texts (Sub-Test 3), and are usually those who read widely and therefore acquired an extensive vocabulary (Sub-Test 4). (p. 22)

The STAR manual contains appropriate reliability information and reports a high correlation with PAT Reading Tests. STAR was reviewed by Greaney (2001) who criticised the test's lack of emphasis on assessing decoding. Such criticism needs to be read in the context of the wider debate of whole language versus phonics; that is, bottom-up and top-down theories of reading. The test was also reviewed by Young (2004) in the *Mental Measurement Yearbook* who described it as an appropriate test for assessing children using the New Zealand Reading Curriculum. Young felt the test would be enhanced with further development through validity studies.

In terms of manageability, STAR takes 45 minutes to administer and takes a similar amount of time to mark as tests such as PAT Reading and asTTle Reading. Like a lot of assessment tools, the utility of STAR is dependent on the use the school makes of the test. Where it is used as a screening device to identify which year 3 and older children require more individual assessment, such as Running Records, rather than assessing all children individually, the test has the potential to improve manageability. However, where schools are assessing children through the use of STAR, PAT Reading or asTTle more than twice a year, manageability becomes a potential issue because of the risk that measurement is eating too much into classroom activities focused more directly on students' learning.

4.8 The New Zealand Curriculum Exemplars

The Exemplars were developed by the Ministry of Education (2003a) and distributed to schools initially from 2003 with progressively more resources being made available. This resource covers all seven curriculum areas. The Exemplars are available to schools in print form, by means of CD-Rom and on the web. They have been described as follows by the Ministry of Education (2004a): "The exemplars are authentic examples of student work. They illustrate learning, achievement and quality in relation to the levels described in the national curriculum statements" (p. 2). The intended purposes of the Exemplars are:

Teachers

- Assess where student is at and set learning goals
- Review learning goals
- Set up peer and self-review
- Collaboratively assess and evaluate progress
- Moderate work
- Review class or school programmes
- Help with class planning

Schools

- Report to parents and caregivers
- Report to the BOT
- School planning and review

(Ministry of Education, 2004. pp. 4-5)

There has been little writing on the use of the Exemplars in schools. Prior to the development, Auckland University was contracted by the Ministry to conduct an international literature review of the use of exemplars (Auckland Uniservices Ltd., 1999).

Poskitt (2002) researched the development of the Exemplars. She found that teachers involved in the development of the Exemplars reported that they were helpful in a number of ways. This included the clarification of expectations about levels of the curriculum as well as expectations about student achievement. Teachers were also positive about how the Exemplars gave them greater understanding of the curriculum especially in areas they did not teach on a daily basis. However the exemplar trial highlighted issues that are common to implementing performance assessment, individual assessment and elements of formative assessment. As was stated in Poskitt, Brown , Goulton and Taylor (2002):

Teachers reported that classroom management issues affected their use of exemplars. Difficulties in using the exemplars in the classroom context included the use of exemplars in multi-level classes, the management of individual teacher-student conversations within a class of thirty students (*what are the other twenty nine students expected to be doing?*), the time involved in annotating individual students' work, management of students simultaneously engaged in a variety of problem-solving tasks, and the use of a range of materials and resources, such as video equipment. Additionally, suggested learning activities in some exemplars raised potential health and safety issues within the classroom, such as groups of young students using Stanley knives.

Comments on teacher work load and time issues featured regularly throughout the consultation and development periods. The teachers expressed concern about how they would incorporate such new developments into their planning and teaching. (pp. 9-10)

Questions concerning the manageability of the Exemplars, like many assessment tools in New Zealand, do not concern the assessment itself but what use the school makes of the procedure. It is questionable whether schools have enough time to undertake individual summative assessment using the Exemplars for all seven essential learning areas. Furthermore, the utility of the collection of such material can also be questioned. As ERO (2000) stated, the common New Zealand approach to teaching science, social studies, technology and health in four to six week blocks means little ongoing use is made of assessment information.

4.9 Numeracy Project

The implementation of the Numeracy Project began in 2000 with a national pilot *Count Me In Too* in 2000 and by 2006 had included 23 000 primary school teachers (Fancy, 2006). In 2002, *The Numeracy Development Project: Policy to Practice* (Higgins, 2002) traced the rationale and policy developments associated with this project. She stated that the Project was a result of a number of concerns about mathematics education in primary schools. This included ERO's 1994 *Mathematics in the New Zealand Curriculum* report that raised concerns about the implementation of the new curriculum and Holton et al. (1996) who highlighted difficulties experienced by teachers. However, the publication of New Zealand's poor results in the TIMSS international survey of achievement in science and mathematics (Garden, 1997) was the biggest driver in this substantial change in teaching practice.

The Numeracy Project has a heavy emphasis on teaching number strategy. The assessment practices associated with the Project are a one-to-one teacher interview termed *the diagnostic, or Numeracy Project Assessment (NumPA)* which involves children demonstrating their stage of development. This assessment takes approximately 15 minutes depending on the students' age and level of development. There are a number of follow up assessments including Global Strategy Stage (GLOSS) which is also a one-to-one interview based on assessing children's progress or understanding after a unit of work. The IKAN tests, which is also part of the numeracy project, are short paper and pencil tests which assess similar information. The results from these assessments are used in a diagnostic/formative manner as well as to group children. The results can be used to a lesser extent for programme evaluation and summative reporting. A revolutionary feature of the assessment for New Zealand schools was that individual school results were recorded electronically on a national data base. In some ways the Numeracy Project can be seen as a window to the future of assessment in that not only is the recording of assessment web-based, the guidelines for its use change rapidly. The Numeracy Project Assessments differ from *MiNZC* in that they are based on stages of development rather than levels of the curriculum.

In terms of manageability, the main numeracy assessment, the full diagnostic, or NumPA, is a one-to-one individual interview which takes approximately 15 minutes per assessment to administer. As such it does present manageability concerns given

the time required to assess all children in the class. However, this comprehensive assessment is supplemented by the less time-consuming GLOSS and IKAN. At the time of writing, the Numeracy Project does not publish guidelines about how often the assessments should be carried out. This decision is left to individual schools and can lead to manageability issues when it is considered that it takes between eleven and twelve hours to carry out a full numeracy assessment for a class.

4.10 asTTle (Assessment Tools for Teaching and Learning)

asTTle has been one of the largest developments in assessment tools in New Zealand. The tests are Literacy and Numeracy assessment tools for students at Levels 2-6 (years 4 to 12) in Written English and Maori. asTTle is an electronic tool with a web-based format in development at the time of this research. asTTle Reading and Numeracy tests are 40 minute, paper and pencil multi-choice tests. The Writing test is based on the analysis of a sample of a child's writing using progress indicators developed for each strand of the writing curriculum. The MoE (2004) described the purpose of asTTle at class level as enabling teachers to:

- diagnose how their students are performing;
- give students specific, focused feedback;
- support students to set realistic learning goals;
- develop and modify class programmes. (p. 12).

It is claimed that at school level the information can be aggregated and used to identify students' needs and evaluate teaching programmes.

4.10.1 Development of asTTle

In the late 1990s, a push for a system of national testing originated from ERO (1998) with support from Nick Smith, Minister of Education for the National Government (MoE, 1998). As part of this development, a contract was given to Auckland University to develop and trial national tests in literacy and numeracy in Maori and English. With the change of government, it was decided to continue the development of the test but to make its use by schools optional.

What is very evident from the documentation of the development of asTTle is that the Project invested far more time and resources in researching and building a comprehensive assessment tool suitable for the New Zealand curriculum in maths and

English than was arguably devoted to researching and building the curriculum itself. The Project's 49 technical reports certainly leave a paper trail of research and rationale that was absent from the original development of the New Zealand Curriculum (MoE, 1993a). The development of asTTle included mapping processes for both the English and Mathematics Curriculum. These curriculum maps provided a basis for item writing, classification and reporting. The curriculum mapping exercise involved teams of experienced teachers and literacy and mathematics experts. As part of the development of the tests it was decided to report student achievement in three sub-levels within each level of the curriculum: basic, proficient and advanced.

As well as being linked to the levels of the New Zealand Curriculum by a process of curriculum mapping, the asTTle items were also classified and structured by use of the Structure of Observed Learning Outcomes (SOLO) cognitive processing taxonomy. SOLO was developed by Biggs and Collis in Australia in the 1980s (cited in Hattie & Brown, 2004) and is based on analysing students' responses to assessment tasks and classifying them into one of four categories: unistructural, multistructural, relational and extended abstract. Unistructural and multistructural responses are described as being concerned with surface features and relational and extended abstract responses are classified as examples of deep features. The asTTle Reading and Mathematics Tests aim to provide a balance of surface and deep items and students' responses are reported in these two dimensions.

Designing and Marking Tests Reading and Mathematics

asTTle allows teachers to design their own test within the asTTle Framework. This means that in reading, teachers can design a test which contains items from any three of the six reading domains: finding information, knowledge, understanding, connections, inference and surface features. Similarly in mathematics: number knowledge, number operations, algebra, geometric knowledge, geometric operations, measurement, statistics and probability. The teacher is also able to choose the difficulty of a test through selecting items from any adjacent three levels of the curriculum: *some level two, a lot level three, a few level four*. The Mathematics and Reading Tests are 40 minute multi-choice tests with the results being marked and entered by the teacher in the computer program.

The results of asTTle tests are reported in a number of ways:

Console The console allows schools to graphically compare their cohorts of students' results against national norms and profile their children's achievement by means of box and whisker graphs. The console also provides data on student achievement in the three curriculum functions that comprise the test as well as information on students' attitudes to the subject and levels of deep and surface thinking.

Individual Learning Pathways This report gives students' results in four quadrants: strengths, achieved, to be achieved and gaps. The Individual Learning Pathways Report also gives students a level in terms of the English or mathematics Curricula which are broken down to three sub-categories: basic, proficient and advanced. The individual pathways report also gives an individual score which can be used for comparison with the mean for the year group. This score can be used to compare the progress of individual students when a subsequent test is undertaken.

New Zealand Comparisons The New Zealand Comparisons Report enables schools to compare their students' achievement with that of a similar decile. The results can also be analysed in terms of gender and both Maori and Pasifika.

Group Learning Pathways This report gives teachers a picture of the strengths and weaknesses of each cohort of children.

What Next The What Next Report directs teachers to appropriate resources to assist in planning and future teaching for a particular cohort of children in terms of the curriculum functions assessed.

Curriculum Levels Report and Tabular Report The Curriculum Levels Report displays students' achievements by levels of the curriculum while the Tabular Report summarises the achievement results for students who undertook the test.

4.10.2 Development of asTTle and Technical Information

As stated earlier, asTTle is supported by a large amount of technical information. Thirty-five of the 49 technical reports are available on the asTTle web site (MoE, 2007c). These reports describe the rationale on which the tests were produced, the development process and in-depth analysis of the students' responses. However, such a large amount of information does mean that it can be difficult to locate the most important elements.

In Technical Report 26, Glasswell and Brown (2003) describe the reliability of marking writing tasks. Sixteen expert teachers with over a day's training were able to achieve 80% agreement on a week long scoring activity. This report also gives an accuracy rate of 70% for untrained markers but provides insufficient information on how many scripts were marked or how this compares with other tests of writing, to enable the implications to be understood and contextualised within schools. It is also difficult to gauge how asTTle compares with other tests of literacy, because the technical information focuses mainly on the analysis of data through item response theory, with traditional methods of estimates of reliability not being reported. This is not a criticism of item response theory – this approach to item analysis and test construction is well recognised as an advance on traditional item analysis methodology. However, the documentation of other forms of technical information, such as correlation studies with other tests with similar purposes, would further strengthen the growing support for asTTle in schools, assuming such evidence was favourable.

The asTTle Project made extensive use of teachers, literacy and numeracy experts and academics, both from New Zealand and overseas, in all aspects of its development. As stated above, the asTTle Project can not only be seen as being closely related to the New Zealand Curriculum Statements in English and Mathematics, but through the high teacher involvement, the curriculum mapping exercises, and the inclusion of over 84 000 children in the norm sample and item responses, it represents a major development of our understanding of the Curriculum. This contributes a great deal to the content validity of the assessment tools. However, while the extensive work and involvement contributes to the validity of asTTle Reading and Mathematics Tests, the fact that there are no correlation studies with similar tests in Mathematics and only a very informal study in reading, poses questions about the validity of the assessment that one would normally expect to be answered in a manual for a conventional test.

In an early examination of asTTle, Darr (2003) acknowledged the tests' ability to allow teachers to identify areas of strength and weakness and provide a measure of how well students are achieving against national standards. However, he expressed a number of concerns. While he considered the reports comprehensive, and well presented they still needed careful interpretation and a good knowledge of the tests. Darr was alarmed over a conclusion that the students had a poor understanding of

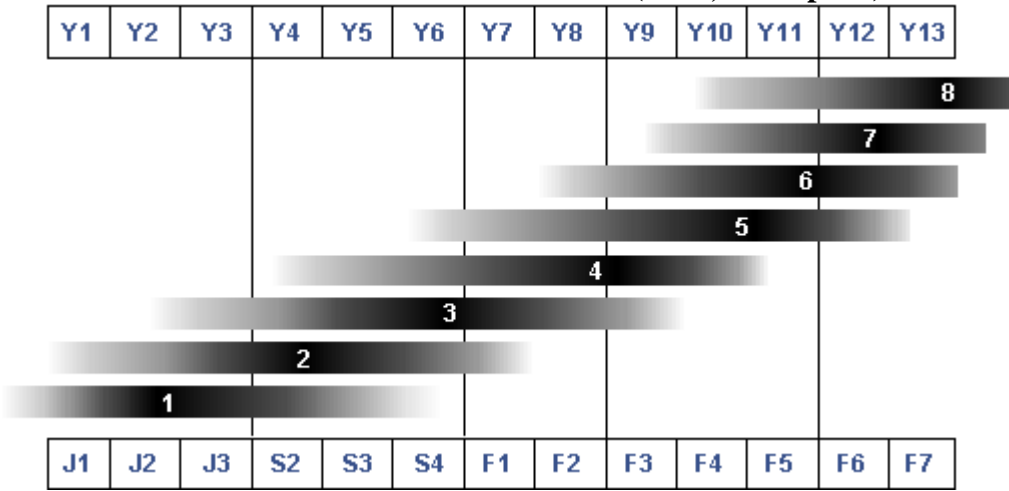
negative numbers based on one question only. This led him to conclude that the report of poor understanding of negative numbers was questionable. Darr also expresses concerns over manageability:

A final caution involves the administration load asTTle requires. Applying asTTle takes time. As well as creating and photocopying the test, the teacher must mark the scripts and enter the marks into the computer. In the busyness of school life, this would probably limit the number of times asTTle would be used, unless help with marking scripts and inputting data was made available to teachers. (p. 14)

In Chapter Three, it was pointed out that the curriculum mapping exercise and the norming procedures involved in asTTle’s development had resulted in a major challenge to the curriculum levels provided by various Ministry documents. The research on asTTle reading indicates the mean achievement for year 8 students being at level 3 whereas the curriculum chart suggests it is at level 4.

Figure 4.1: Levels of the New Zealand Curriculum and Year Levels

Mathematics in the New Zealand Curriculum (MoE, 1992. p. 17)



Furthermore, asTTle’s views on curriculum levels brings into question ERO’s criticism of some schools over 15 years for not operating a best fit assessment system for levels of the curriculum. If the curriculum developers had it seriously wrong in the first place, why should individual schools have been expected to get it right? Given that in the development of asTTle thousands of hours were invested to come to the conclusions that the levels of achievement were inaccurate in the original curriculum documents for mathematics and English, ERO could not possibly expect schools to assess against the levels in science, social studies, and technology etc. We must therefore be cautious about interpreting students’ assessment in the curriculum areas where research has yet to support the year levels shown above in Figure 4.1.

The above discussion illustrates that while levels of the New Zealand curriculum would have provided some guidance for teachers for planning purposes, they were totally inappropriate for assessment. This makes the brief for the development of asTTle extremely challenging as it is very difficult to base an assessment system on something that is itself lacking in research, logic, consistency or a well-developed understanding of how children learn (Elley, 1993; 2004). The asTTle project has certainly left a far more substantial research and development paper trail than was the case with the original development of the curriculum documents.

4.10.3 The Manageability of asTTle

In terms of manageability, asTTle Mathematics and Reading take a similar time to administer and mark to that of other norm referenced multi-choice tests such as STAR and PAT. The test takes 45 minutes and a conservative estimate of 2 1/2 hours to mark the test and load the data is, as stated above, similar to other standardised tests. As has also been stated at other points in this thesis, the New Zealand approach to assessment means that it is up to individual schools to decide whether they want to use the tests and how often they wish to administer them. The testing regime can become unmanageable if schools decide to undertake a large amount of standardised testing. There is no information as to whether or not this is the case, or even what could be considered excessive standardised testing. However, the New Zealand approach, emphasising the self-management of assessment, does leave schools vulnerable to such problems.

The manageability of asTTle writing does present more challenges. The procedures appear to be complex and considerable training and moderation is required in order for schools to produce credible data. The fact that asTTle Writing cannot be used with children younger than year 4 also presents manageability problems. It is common for New Zealand schools to have composite classes (two or more year levels in the same room). Schools with composite year 3-4 classes have to use two forms of writing assessment within the same classroom.

4.11 Recent Developments

Initiatives by the Ministry of Education since 2000, termed the *National Assessment Strategy* (Chamberlain, 2000), have aimed at supporting teachers and addressing

current issues in assessment. The Assessment Community on the TKI website (<http://www.tki.org.nz/r/assessment/>) provides comprehensive support for teachers. This includes access to the main assessment tools (ARBs, The Exemplars, and asTTle), school case studies, and professional reading. This site and recent MoE publications give guidance to teachers on formative assessment and evidence-based teaching, the major theme being promoted by the MoE. These changes also reflect an emphasis on improving the quality of teaching in New Zealand as a major strategy in addressing under achievement (Alton-Lee, 2003; MoE, 2004a, 2004b; Thrupp, 2004, 2005).

4.12 Mathematics Assessment

While this chapter has focused on individual assessment tools a short discussion on assessment by subject illustrates some issues that schools have to face when selecting assessment tools. Table 4.2 lists the mathematics assessment tools available for use in New Zealand schools.

Table 4.2: Mathematics Assessment Tools in use in New Zealand Schools

Tool	Level	Reporting Information
School Entry Assessment	New Entrants (5 year olds in their first six weeks of school)	
Numeracy Project Assessment (1) Full diagnostic (2) Gloss (3) Information sheets	Years 1-8	Stages in Numeracy Project
The Exemplars	Levels 1-5 (all levels of primary school)	Levels of the Mathematics Curriculum broken down to three sublevel
Assessment Resource Banks	Levels 2-5 years 3-4 to 8	Levels of the Mathematics Curriculum
PAT Mathematics	years 4-8	Class and age stanines
NEMP	years 4-8 Can be used at all years between 4 and 8	National results at Years 4 and 8
AsTTLe	years 4-8	Level of the New Zealand Curriculum broken down in three sub levels: Basic, Proficient, Advanced Mean Year level Scores
Learning Objects (NZMaths website)	Levels 1-5 New Mathematics Curriculum	Levels of the NZ Curriculum
TIMSS Mathematics	Year Five	Compared with age and year groups
School developed test	All levels	

Note that this table excludes some assessments used by a small number of schools or small groups within schools such as the BSM checkpoints (these were common in junior classes before the Numeracy Project), and assessments aimed at more able students such as the Australian Tests and Otago Problem Solving.

This means that in New Zealand primary schools, excluding school developed procedures, there are eight major mathematics assessment tools available for use at year 4 or above. If we include the latest PAT Mathematics test in this eight, all have been developed since 1995; the most commonly used assessments, the Numeracy

Project and asTTle, only became substantially available since 2000-2002. What is more, the results of the different assessments are reported in different ways and while four tools (asTTle, ARB, The Exemplars and Learning Objects) report by levels of the curriculum, there are three variations on how this is done.

Such an extensive number of assessment systems have a number of implications for the education system. It means that when ERO reviews a school they could be faced with a large number of different formal assessment procedures in Mathematics. This has implications for their ability to gather efficiently a picture on achievement in individual schools and make comparisons with similar schools. It also has implications for the professional development of teachers. Rather than have teachers become experts at carrying out and analysing assessment in two assessment procedures such as the Numeracy Project and asTTle it means that professional development providers have to design programmes for all assessment procedures. It also means that when teachers move school they may encounter an entirely new assessment regime that they need to master. In a similar way, the existence of multiple assessment procedures means that sharing results with parents can also be confusing when a child moves schools. This proliferation of assessment procedures could be questioned in terms of a prudent use of vote education. NEMP, ARBs, the Exemplars, Learning Objects, asTTle and the Numeracy Project were funded directly by the Ministry of Education while the large part of PAT development comes from sales to schools paid for by their operation grant.

4.13 Assessment in the Non-Core Curriculum Areas: the Visual Arts

While the literature is not extensive, arguably there has been enough published information to gain an understanding of procedures for assessment in numeracy and literacy in New Zealand schools. However, assessment in other areas of the curriculum has been subject to far less scrutiny. It is harder to gain an understanding of what is happening in the other five areas of the curriculum: science, social studies, the arts, technology and health and physical education. Space does not allow a close scrutiny of all these areas. However, an examination of assessment activity in the visual arts gives an indication of how schools conduct assessment in areas other than numeracy and literacy.

Arts in the New Zealand Curriculum (2000) was the last of the curriculum documents to be implemented. In terms of assessment, both the New Zealand Curriculum Exemplars and NEMP provide assessment tools in Art and there are examples of assessment procedures in the Art page of the TKI website (MoE, 2007).

As stated above, assessment in art has been subject to very little discussion in the New Zealand education community. ERO documents provide the most substantial writing on issues in art education. The Education Review Office's publication is based on their interpretation of the 1993 NAGs.

Through the principal and staff, boards of trustees are required to:

- I. Implement learning programmes based on the underlying principles, stated essential learning areas and skills, and the national achievement objectives;
- II. Monitor student learning progress against the national achievement objectives; and
- III. Assess student achievement, maintain individual records, and report on student progress

(Education Gazette, 30 April 1993, p. 3).

In 1995, ERO outlined expectations for schools in the assessment of art in their Report *Arts in the New Zealand Curriculum*:

Boards of trustees should be informed about the learning experiences provided in arts programmes in their school. Trustees and families should expect to see evidence of student achievement in all three components of the curriculum statements in art and music as well as in other areas of arts education. Boards of trustees should expect the principal to report on student achievement in arts education. (p. 6)

In their 1999 report, *Student Assessment: Practices in Primary Schools*, ERO examined assessment in literacy, science, numeracy and art in 118 primary schools. The report was published shortly after the *Arts in the New Zealand Curriculum* was published and before schools were expected to have implemented the document. The study found assessment procedures were poorly developed in the schools included in the study. "Assessment procedures varied a great deal within schools. Even where schools had developed some form of systematic data gathering teachers in most of these schools depended on subjective judgements very loosely guided by learning

objectives or statements of skills. These were open to a considerable range of interpretations” (p. 29).

The report identified that only 14% of the schools were using assessment activities to help improve their art programmes. Four out of 118 schools were making use of NEMP Art assessment tasks. The Report acknowledged that the lack of attention given to the assessment of art in primary school may arise from manageability issues:

The lack of systematic assessment in art may be, at least in part, a result of conscious decisions to give greater priority to the literacy and numeracy areas of the curriculum. It may also be a result of the lack of any national or commercially available set of procedures that teachers could adopt. (p. 27)

The Report concluded:

The current state of assessment in art at primary school level, and the fact that records of achievement in art were for many years not required for students in the junior classes of primary schools, raises a number of questions:

- What sort of information would be useful to teachers, students and parents?
- How much of this information should be recorded?
- What priority should be given to the use of teacher time for this task?
- What sort of system-wide assistance should be given to teachers through the development of assessment tools and exemplars?

Teachers and perhaps parents also, need to agree on the fundamental purpose or purposes for which assessment information about art is needed, on how the information could be obtained and on how it could be used to fulfil those purposes. (p. 29)

In their 1995 and 1999 reports, it is obvious that ERO placed emphasis on a *balanced curriculum* and subsequent high expectations for assessment in such areas as art. In 1999 (MoE, 1999b) changes were made to the NEG's to “give greater priority to numeracy and literacy.” (p. 3). There was, therefore, an implication that less emphasis would be given to assessing other areas of the curriculum, including art, especially in the beginning years of school. However, it is unclear from Ministry of Education guidelines whether this means a lowering of expectation in the assessment of Art. As stated in *Sharpening the Focus* (MoE, 2000b):

Are schools still required to provide a "balanced" curriculum?

The revised NAGs make it clear that priority is to be given to student

achievement in literacy and numeracy, especially in the early years. Beyond literacy and numeracy, the Ministry considers that the balance of essential learning areas and essential skills is a matter for each school to determine within the framework provided by the national curriculum statements. (p. 23)

In 2004, ERO noted in their report on the quality of teaching Visual Arts in years 4 and 8 that there were issues in assessment. In particular, the use of assessment information was not always or rarely effective in over half (51 percent) of the 213 schools in the sample.

Art has been assessed by NEMP in 1995, 1999 and 2003. The 1999 Forum Report recorded that children enjoyed art with it being the favourite school activity in year 4 and the second most popular in year 8. However, the writers reported concerns that there was little change in children's ability to create or discuss art work from year 4 to Year 8 with the exception of drawing. The 2003 report, which had followed the implementation of the Art Curriculum and subsequent professional development, the Forum Report noted that:

Few students were able to use paint with confidence to make personal imaginative statements.

In responding to artworks, most students had difficulty going beyond description of recognisable subject matter, to elaborate, interpret, construct, question, argue or analyse their initial observations.

Compared with 1999, year 8 students in 2003 reported fewer opportunities to talk about art. Year 4 students have greater access.

In conclusion, not a great deal is known about assessment in art in New Zealand primary schools. School leaders and teachers may conclude that ERO's expectations about art assessment are high and guidance from the MoE is vague. It would be interesting to investigate how widely and in what capacity the Art Exemplars are being used.

4.14 Summary of New Zealand Assessment Tools

This chapter examined the major assessment tools and tests that are currently used in New Zealand schools. There was an emphasis on those tools created since the reforms such as the ARBs, NEMP Tasks, asTTle, The New Zealand Curriculum Exemplars and the Numeracy Project. The chapter concluded with a discussion on the assessment of art as an illustration of assessment of the non-core curriculum. Developments in

assessment practice in New Zealand schools over the last fifteen years have seen a change from the early nineties when schools were directed to assess a new curriculum without assessment tools available to carry out these tasks, to the situation in the mid years of the new millennium where there was a proliferation of assessment tools.

CHAPTER FIVE: METHODOLOGY

5.1 Introduction to Methodology

This chapter outlines the methodology used in this research. It discusses the rationale for the choice of the interpretivist/constructivist research paradigm and case study methodology. Issues such as validity/trustworthiness, transferability and ethics are explored and their application to this research is detailed. The use of the major research tool, semi-structured interviews, is also discussed and the methods used for the analysis of the research data are explained.

A basic principle of any research is that the methodology should be appropriate for the context of the study and the research questions it seeks to answer. Therefore it is appropriate to begin this chapter by restating the research questions.

1. What were the major influences and challenges to assessment in New Zealand primary schools since 1990?
2. How did three New Zealand primary schools establish assessment and reporting systems during the period 2003-2007 that were manageable and maintained a focus on learning?
3. To what extent do the staff of three case study schools find the most recent changes in assessment practice/tools manageable and provide useful information?

5.2 A Constructivist World View

While the research questions consider primary schools, a central basis for this study is that schools are composed of people: teachers, pupils, principals, support staff, BOT members and parents. It is the interaction of people, in particular management staff and classroom teachers with assessment and reporting procedures that is the core of this investigation. A further central element to this research is that it is not focused as much on concrete documents, such as assessment policies, reports to BOT, ERO

reports and test scores, as on the way these documents are perceived by the people involved in their administration and implementation in the three schools. As such, this study is based firmly on a constructivist world view.

A constructivist world view takes the approach that it is impossible to know the human world in a truly objective sense. Constructivists hold the view that human beings have evolved the capacity to interpret and construct reality (Patton, 2002). The human world of schools, families, organisations can only be understood from the construction people make of their lives, their actions and the actions of others. In the words of Patton “constructivists study the multiple realities constructed by people and the implications of those constructions for their lives and their interaction with others.” (p. 96)

In relation to this research, the fundamental issue is not whether the school or the teacher is using the STAR reading tests or asTTle Mathematics but whether perceptions indicate the test is manageable and effective, and makes a valuable contribution to their teaching and reporting to parents. It is from seeking to understand the perspectives of those involved individually and collectively that we move toward a fuller understanding of what is happening with assessment in our schools, or more correctly the three schools under investigation. This is not to say that other paradigms and methodologies would not also be appropriate in seeking to gain a deep understanding of assessment in New Zealand primary schools; the important point is that the approach taken has been chosen because of its fit to the questions being asked.

5.3 Approaches to Research

In broad terms research inquiry in education tends to follow one or a combination of two major distinct approaches: the normative or positivist paradigm which is based on traditional scientific method and quantitative data; and the interpretivist paradigm which is associated with methods that involve the collection and analysis of qualitative data (Cohen, Manion, & Morrison, 2000).

Research on assessment has a strong association with the positivist paradigm. In the New Zealand context this is most often associated with the development of norm referenced assessment tests: asTTle (Glasswell & Brown, 2003; Meager-Lundberg & Brown, 2001), PAT (Reid & Elley, 1991) and STAR (Elley, 1985, 1997). Such

research has an emphasis on numbers and measurement activities associated with psychometric assessment which, until the last 20 years, has dominated discussion and research in educational assessment. This approach to measurement has resulted in the development of a large number of tests, rating scales and other assessment techniques that have provided data for experimental and other forms of normative research. It has also resulted in the emphasis that assessment developers have given to the concepts of reliability and validity and the statistical techniques associated with such concepts. While a quantitative approach to assessment has helped promote understanding around the concepts of reliability and validity, it has shed little light on the important aspect of the manageability of assessment. Although it may be possible to identify through various procedures and experimentation some of the dimensions of manageability and then construct quantitative items to measure these dimensions, it is hard to see how such methodology would be more helpful in understanding the way teachers and schools react to and engage with new initiatives in the curriculum and assessment, than through research which is based on qualitative data. The essential point is that the nature of the reaction or engagement of teachers to the manageability of assessment is better understood by allowing themes to emerge from the experiences, or more specifically, the perceptions of these experiences as described by teachers themselves, rather than the researcher attempting to quantify such experiences or perceptions on the basis of the researcher's pre-coded categories. This point is consistent with a theory of research that emphasises constructivist approach to making sense of people's experiences.

5.4 Positivistic Paradigm

Maykut and Morehouse (1994) define a *paradigm* as a set of overarching and interconnected assumptions about the nature of reality. Cohen, Manion and Morrison (2000) outline the major assumptions that guide scientists who adhere to the positivist paradigm:

Determinism: events in the natural and social world are caused by other events and circumstances. There is some order to these series of causes which can be discovered and explained. The discovery of these causes allows scientists to formulate rules that allow the prediction of future behaviour.

Empiricism: the view that knowledge can only be based on what can be observed and measured. This observation allows scientists to test hypothesis or theories.

Parsimony: This principle states that any explanation for phenomena should be stated in the most economic manner possible.

Generality: Findings about specific concrete events should be able to be generalised to the world at large. (p. 10)

The positivist paradigm is thus concerned with the traditional scientific method which has a number of key characteristics. It is based on objective observation. It is concerned with generating theory, testing hypotheses, and using standard methods that are capable of repetition by other researchers. A key feature of a traditional scientific approach is the central role statistical analysis plays in research. A statistical approach is embedded within the traditional scientific method and the application of probability theory is a major element in positivist methodology.

While the positivist paradigm and the scientific method have been successful in expanding knowledge in the physical and biological sciences, the adoption of this methodology has been subject to criticism as being inappropriate for expanding understanding in the social sciences. As Cohen et al. (2000) state, “The precise target of the anti-positivist’s attack has been on science’s mechanistic reductionist view of nature which, by definition, excludes notions of choice, freedom, individuality and moral responsibility” (p. 17). In other words, an approach that does not take into account the ability of people to act as individuals, and to behave in unpredictable ways through a variety of motivations and rationales, is limited in attempting to shed light on the complex social world.

A further criticism of the use of the scientific method in attempting to inform better practice in education is that, despite thousands of hours of research and countless studies, this approach has contributed much less of value to the education community than one would expect from the effort and resources put in. To quote Cohen et al. (2000) again:

The findings of positivist social science are often said to be so banal and trivial that they are of little consequence to those for whom they are intended, namely teachers, social workers, counsellors, personal managers and the like. The more effort, it seems, that the researchers put into their scientific experimentations in laboratory by restricting, simplifying and controlling

variables, the more likely they are to end up with a pruned synthetic version of the whole, a constructed play of puppets in a restricted environment.

(p. 19)

A number of terms are often used interchangeably in discussion of research methods such as positivism and quantitative research. There can be subtle differences between the two. In its purest form positivism is concerned with the classical scientific method and research that tests theory. The difference between quantitative research and qualitative research is more about what is measured objectively and counted (quantitative) and what may be observed interpreted (qualitative), or as Strauss and Corbin (1998) state:

By the term qualitative research we mean any type of research that produces findings not arrived at by statistical procedures or other means of quantification. It can refer to research about person's lives, lived experiences, behaviours, emotions and feelings as well as about organisational functioning, social movements and cultural phenomena and interactions between nations. Some of the data may be quantified as with census or background information about the person or persons being studied, but the bulk of the analysis is interpretative. (p. 11)

However, it should be noted that a number of New Zealand studies that have shed some light on the implementation of assessment in this country have included quantitative data. For example, Croft, Stafford and Mappa (2001) and Dunn and Marston (2003) examined the use of assessment procedures in New Zealand schools; these are studies which included measures of what assessment tools were being used and how effective they were rated by teachers. Timperley (2003a) measured the amount of time being used for taking Running Records in year 4 – 6 classes and concluded that 20% of reading instruction time was being used for Running Records. Wylie (1997) used the analysis of extensive survey data to measure the impact of the education reforms on teachers, principals and BOT members.

5.5 The Interpretivist Paradigm and Qualitative Research

In attempting to expand understanding of the complex social worlds of classrooms and schools many education researchers have adopted an interpretivist approach. Once again there are a variety of terms associated with the interpretive paradigm. They include qualitative research methodology which will be expanded on in detail below. Denzin and Lincoln (1993) acknowledge that attempting to define qualitative research is complex and challenging as the research community's understanding of

the concept has changed a great deal over the last hundred years. In addition, the term has been influenced by the whole field of social science and more recently a number of disciplines in the humanities. However, with these cautions in place they offer the following definition:

Qualitative research is a situated activity that locates the observer in the world. It consists of a set of interpretive, material practices that make the world visible. These practices transform the world. They turn the world into a series of representations including field notes, interviews, conversations, photographs. At this level, qualitative research involves an interpretive, naturalist approach to the world. This means qualitative researchers study things in natural settings, attempting to make sense of, or to interpret, phenomena in terms of meanings people bring to them. (p. 5)

The interpretivist paradigm and qualitative research can be explained in a number of ways by contrasting it with positivism. Positivism and its accompanying traditional scientific method stress the objectivity of the researcher. The interpretivist paradigm acknowledges the major role of the researcher in their interaction with the subject of the study. Whereas a traditional scientific approach in the study of human behaviour, will focus on observable and measurable external behaviour, the interpretivist approach is concerned with a subject's internal explanations and rationale for their behaviour and the behaviour of others. As such, this approach is concerned both with observable behaviour and how subjects experience or rationalise their actions.

Positivism is concerned with generating and testing theory. The role of research and experiment is to test theory; to accept or disprove an hypothesis. As such theory is tested in research in an attempt to develop a more general theory that accounts for a wider spectrum of human behaviour (Cohen et al., 2000). In an interpretive approach, theory is generated from the research rather than preceding it. While positivism is concerned with developing theories that can be generalised to a wide range of situations, interpretive research and its associated qualitative methodologies are concerned with deep understanding of very specific phenomena that cannot usually be generalised to other situations.

Denzin and Lincoln (2000) also contrasted the difference between a positivist approach and the interpretivist paradigm in discussing the approaches of qualitative and quantitative researchers:

Qualitative researchers stress the socially constructed nature of reality, the intimate relationship between the researcher and what is studied and the

situational constraints that shape inquiry. They seek to answer questions that stress how social experience is created and given meaning. In contrast quantitative studies emphasize the measurement and analysis of causal relationships between variables, not processes. Proponents of such studies claim their work is done from within a value-free framework. (p. 8)

There have been a number of criticisms made of the interpretivist paradigm. Rex, cited in Cohen et al. (2000), makes the point that while knowledge of the subjects' intentions is important in understanding their actions, a more objective view is necessary if we are going to take credence from any research findings. While it is helpful to interview participants and their understanding of events and actions, they may recall such events inaccurately.

While there have been well articulated criticisms of both the positivist and interpretivist paradigms, both approaches, if well selected and appropriately applied, have a lot to offer education research. The key point is that the research methodology must be the one that is the most appropriate for the particular study. As Silverman (2005) states: "Any good researcher knows the choice of method should not be predetermined. Rather you should choose a method that is appropriate to what you are to find out." (p. 11)

5.6 Pragmatism - Mixed Methods Approaches

Robson (2002) argues that what he terms a pragmatic approach to methodology is most appropriate for those working in applied fields: "use whatever philosophical or methodological approach works best for a particular research problem at issue. This leads to mixed method studies where both quantitative and qualitative approaches are adopted" (p. 43). He goes on to state that such a pragmatic approach has a philosophical base in American history, as evidenced in the work of James and Dewey. For pragmatists "truth" is closely aligned with what works. Creswell and Plano-Clark (2007) also embrace pragmatism with their argument for a mixed method research; like Robson, they advocate for research that works rather than strict adherence to a particular paradigm or methodology.

The argument for pragmatism and mixed method research has relevance for this study. Although the research mainly uses a qualitative methodology the investigation records which assessment tools are used in the schools and how often they are applied. This recording of assessment tools and their frequency of use provides some of the context for understanding the participants' perception of manageability and

utility. Without some measurement of how many times a year a school undertakes, for example, Running Records, asTTle and STAR tests, it would be difficult to put into context teachers' comments about manageability of the tools in question. However, while such recording tends towards the quantitative, the dominant approach used in this study is qualitative which permits the research to focus on the understanding that principals, management staff and teachers have of manageability and utility issues in assessment. It attempts to give a deep insight into these issues in the three case study schools.

5.7 Selection of Methodology: Case Study

This study sought to understand assessment and reporting systems in three schools. In that schools are organisations located in specific places and with a relatively clear membership (students, teachers, management, support staff, volunteers, parents and BOT) they can be considered bounded systems. The bounded nature of a school makes a case study an appropriate approach in undertaking research. Cohen et al. (2000) described case studies as being set in a context of time, place, organizational and institutional context that allow boundaries to be drawn around the case. They also state that case studies can be defined by the characteristics and roles of participants in the case.

In attempting to define case study research, Merriam (2001) placed emphasis on the singularity of the unit being investigated: "a qualitative case study is an intensive, holistic description and analysis of a single instance, phenomenon or social unit" (p. 27).

Bassey (1999) outlines a comprehensive definition of the educational case study which provides not only guidance but some lofty goals to aim for.

An education case is an empirical inquiry which is:

- conducted within a localised boundary of space and time (i.e. a singularity);
- into interesting aspects of education activity, or programme, or institution, or systems;
- mainly in its natural context and within an ethic of respect for persons;
- in order to inform the judgments and decisions of practitioners or policy-makers; or of theoreticians who are working for these ends;
- in such a way that sufficient data are collected for the researcher to be able

- a) to explore significant features of the case,
- b) to create plausible interpretations of what is found,
- c) to test the trustworthiness of these interpretations,
- d) to construct a worthwhile argument or story,
- e) to relate the argument or story to any relevant research in the literature,
- f) to convey convincingly to an audience this argument or story,
- g) to provide an audit trail by which other researchers may validate or challenge the findings, or construct alternative arguments. (p. 58)

Bassey (1999) goes on to describe three types of case study: theory-seeking or theory-testing case studies; story-telling and picture-drawing case studies; and evaluative case studies. He describes theory-seeking or theory-testing case studies as “particular studies of general issues aiming to lead to fuzzy propositions (more tentative) or fuzzy generalisations (less tentative) and conveying their context and the evidence leading to them to interested audiences” (p. 58). Story-telling and picture-drawing case studies focus on describing and analysing educational events, projects and programmes. Evaluative case studies involve analysis of educational programmes, projects, systems or events to determine their worthwhileness. This research contains elements of all three types of case study. It explores and theorises on the concepts of manageability and utility within the context of assessment practice in the primary school. It attempts to draw a picture on the way three schools have adapted to changes based on the reforms of the school curriculum and assessment, through analysing developments as told by participants and where the developments have both historical and current context. The research also engages in evaluation; the focus of the evaluations includes analysis of documents, the research literature and participants’ experience of different assessment tools and systems.

Stake (2000) also classified case studies in three ways: intrinsic case studies, instrumental case studies and collective case studies. An intrinsic case study focuses on the “case” itself – the case rather than a particular phenomenon or event, is the main interest. In contrast an instrumental case study establishes a bounded case in order to examine a broader issue, or theory, or phenomenon. A collective case study involves the use of a number of bounded cases to examine a phenomenon in different environments. Clearly, the important point for this study is that no two schools will have adopted exactly the same assessment systems, practices and tools. For each, the experience of manageability may have been different, but it might also have contained some overlap.

This research would fit into Stake's (2000) description of an instrumental case study in that it is not the schools themselves as functioning units that are being examined, rather than the experience of staff within the schools in relation to the manageability and utility of the assessment tools and practices that they have engaged with. As Stake notes:

I call it an instrumental case study if a particular case is examined to provide insight into an issue or to redraw a generalization. The case is of secondary interest, it plays a supportive role, and it facilitates our understanding of something else. The case is still looked into in depth, its contexts scrutinized, its ordinary activities detailed, but all because this helps the researcher examine the external interest. (p. 437)

Patton (2002) states that the term a case study can be used to describe a product of a research process. He sees a case study as:

... a specific, way of collecting, organising and analysing data; in that sense it represents an analysis process. The purpose is to gather comprehensive, systematic in-depth information about each case of interest. The analysis process results in a product: a case study. Thus the term case study can refer to either the process of the analysis or the product of the analysis or both.

(p. 446)

In this research, the case study schools are the context used to explore the major issue of the manageability and utility of assessment for teachers and school leaders working in the three schools. In this sense, this study potentially could have elements of other methodologies within the broader qualitative research framework, but they have not been employed in any deliberate sense so do not warrant any claim about their use. For example, an interviewing approach more characteristic of phenomenology could have been used to obtain the experiences of participants in relation to the phenomenon of the study "the manageability and utility of assessment". However, as will be seen shortly, the interview schedules used for this study were quite directive because of the researcher's desire to obtain data on specific assessment practices and developments.

In summary, the research uses case study methodology where manageability and utility are the phenomena under investigation and the three cases (schools) are bounded systems located in time and place with defined membership. The research methods of interviews and document analysis are typical of case study research. However, these techniques are also common to many other forms of qualitative research. The study also contains some quantitative elements in that the tests and

other assessment procedures that each school administers, and their frequency of use, are recorded to provide a context to understand the teachers' interview responses. The dominant aspect of the research is the use of semi-structured interviews which seek to provide the perspective on assessment of teachers and school leaders. In both the method of data analysis and the search for a cohesive understanding of assessment in the research site schools, data are analysed in a format of thick context descriptions that are characteristic of qualitative research methods (Creswell, 1998; Denzin, 1998; Hauberman & Miles, 2002; Mertens, 1998; Merriman, 1998). Mertens describes this as "extensive and careful description of time, place, and culture" (p. 183). When such an approach is carried out effectively it allows an interested reader to decide whether these thick descriptions transfer to other contexts.

5.8 Semi-structured Interviews

As stated above, the use of semi-structured interviews was the dominant research technique used in this research. While the study also included document examination in terms of school policies and procedures in assessment and reporting, semi-structured interviews of school leaders and teachers made the major contribution to the research's conclusions. A number of the conclusions drawn from the semi-structured interviews were subsequently examined in a series of three focus groups interviews of principals. This process led to some modification of the conclusions developed from the semi-structured interviews and documentation examination in the three case study schools.

Qualitative interviews can be seen as being on continuum from structured interviews with a predetermined format and closed questions to unstructured open questions. Scott and Morrison (2006) describe this continuum:

At a general level, interviews sit in various positions upon a continuum of qualitative-quantitative approaches to research. At one 'standardised' end are highly structured interview surveys that pay close attention to the task of collecting large amounts of data, in as focused a way as possible, through the use of proforma like ringing codes, the use of numerical values, tick boxes, and so on. Here, as May (1993:93) puts it the interviewer attempts to control and 'teach' interviewees to 'reply in accordance with interview schedules'. At the other end, there are semi and unstructured interviews that encourage interviewees to respond open-endedly and 'to answer a question on his or her own terms.' (p. 133)

Semi-structured interviews can be described in a number of ways. Gall, Gall and Borg (2007) describe the process as one that “involves asking a series of structured questions and then probing more deeply with open-form questions to obtain additional information” (p. 246).

The interviews used in this research align most closely to a semi-structured interview format. While the interview questions were mostly open-ended in style, they were directive in their focus because of the researcher’s desire to obtain participants’ reactions to specific developments in assessment. For example, the principal, senior management and teachers who were interviewed in each school were asked to identify which assessment tools they used, whether they found them manageable and whether they found them useful. Probe questions were then asked to explore further. The interview schedules are contained in Appendices A and B. There was a slightly different format for the interviews for the principals and other senior management, and scale A teachers.

The rationale for the choice of interview format as the main source of data gathering was to obtain a rich source of information about the participants’ perception of assessment practices in their school.

5.9 Document Examination

As part of this research there was an element of document examination. These documents fell into two categories. The first included such items as the three schools’ written policies and procedures relating to assessment, and examples of portfolios and school reports. The second category was the individual school’s ERO reports. Both these sources of documents could be described as inadvertent sources (Duffy, 1999). Duffy states that: “Inadvertent sources are used by the researcher for some purpose other than that for which they were originally intended. They are produced by the processes of local and central government and from the everyday working of the education system.” (p. 109).

5.10 Focus Groups

In addition to the use of semi-structured interviews and documentation examination the research also used a focus group of principals to examine conclusions from the literature and case studies. Cohen, Manion and Morrison (2000) define focus groups

as “contrived settings, bringing together a specifically chosen sector of the population to discuss a particular given theme or topic, where interaction with the group leads to data and outcomes” (p. 288).

Krueger (1998) describes a number of uses for focus groups including generating and evaluating data from different sub-groups of a population and gathering feedback from previous studies (p. 6). It is these elements that focus groups were used in this research. The semi-structured interviews and literature search led to the development of two frameworks which sought to explain the development of assessment in New Zealand during the period being investigated. These frameworks, which have been titled the *three waves of assessment* and *four phases of assessment change*, were evaluated by three focus groups of principals in order to give the structures greater validity.

The focus group participants were also asked to comment on the section at the beginning of Chapter Three that covers assessment in New Zealand primary schools in the 1980s. There was little literature that documented this period and this section is largely based on the researcher’s own memory of teaching in that period. The focus group confirmed that this section captured their own recollections of assessment during that period.

The participants in the focus groups were all current principals in the Wellington region who had either been principals or in leadership positions since the early 1990s.

5.11 Conducting the Research

5.11.1 Research Sample

The study required both a deep investigation and one that took account of the rapidly changing assessment landscape in New Zealand. For this reason the investigation was restricted to three case study schools. It was believed that three schools would provide enough variety to give the study more credibility than if the investigation was restricted to researching one school. While each case study has been analysed as a single entity it was important to study the implementation of assessment in more than one context; as noted previously, no two schools will have adopted exactly the same systems, practices and tools. While there is significant limitation on how much the

findings from case study research can be generalized to other contexts, research in a single school would narrow the utility of the findings even further.

5.11.2 Selection of Case Study Schools

In order for the research to be itself manageable the selection of the schools was restricted to those in the greater Wellington region. However, this created some challenges. In the last 17 years I have been principal of three Wellington schools. I know many of the principals well and a number of them are close friends. In the interest of validity it is important that the schools in the study did not include those where there is a close relationship between myself and the principal of the case study school.

It was also important that at least one, but not all of the schools, had participated in the MoE's Assess to Learn (ATOL) professional development contracts. This was important as ATOL is the only substantial professional development programme in assessment offered to schools. Without including a school that had undertaken ATOL, it could be argued any conclusions reached could be due to the absence of systematic professional development in assessment.

It was deemed important also that the three schools were geographically spread. While they are independent entities, schools do collaborate in clusters and there is more likelihood of neighbouring schools adopting similar approaches to assessment. For this reason the schools came from three of the four regions of greater Wellington: Wellington City, the Hutt Valley and the Porirua Basin. There was no school from the Kapiti Coast.

It was also felt that the schools should represent a variety of deciles (the New Zealand system of socio-economic rating of schools for differential funding, with decile 1 schools having the lowest socio-economic rating). The three schools comprise decile 2, 7 and 10. It was also decided that the schools should cover both structures common to New Zealand primary schools; full primaries (year 1 to 8) and contributing schools (year 1 to 6)

5.11.3 Recruitment of Schools

Three schools were initially selected on the criteria listed above and the principal was approached by phone. The research was explained to them and a request was made for them to participate. The three schools agreed to participate and a formal request was made to their Board of Trustees. For the purposes of this research, the schools were given the names Endeavour, Kupe and Tasman.

5.11.4 Interview Procedures

The three principals were interviewed in term four (November) 2006 and the tapes transcribed in the January holidays. This was followed with three more participants being interviewed in February 2007. The final five participants were interviewed in May 2007 when I commenced my period of 20 weeks study leave from my principal's position. This meant that the interviews for Kupe and Tasman stretched over a six month period.

The duration of the interviews was from 1 hour to 1.5 hours and were recorded and later transcribed. The interview format began with an explanation of my understanding of the concepts of manageability and utility. This was followed by the respondent giving a description of the major assessment activities in the school. The respondents were then asked to describe what was going well in assessment, what was challenging and what aspects of assessment were in need of change.

The participants were also asked about any ongoing or previous debate among staff regarding assessment issues. This was also followed by a section that sought to provide an understanding of the respondents' assessment background in terms of relevant courses, reading and people who had influenced their understanding in this area. The interview then questioned the respondents about specific assessment practices from the perspective of manageability and utility. A number of the follow-up questions attempted to gain more background on historical assessment practices used during the period 1993 to 2000 both in the respondents' present school or other schools that they had taught at if this was appropriate.

The interviews were held at the participants' schools in the principal's office, teacher's classroom or staff room depending on what was most appropriate. The choice of location was decided by the participants to provide more rapport between

myself and the person being interviewed. Interviews were largely carried out after school or during teacher release time.

The major criticism of the use of interviews in the literature is the time required to transcribe the data. I undertook all transcriptions myself. Most interviews took six to seven hours to transcribe. This was considered a time-consuming but necessary procedure.

5.11.5 Documentation Examination in the Case Study Schools

As stated above, in addition to the interviews assessment documentation at the three schools was examined. This included reports to the BOT, Planning and Reporting data for the Ministry of Education, ERO Reports, school assessment policies and procedures, student profiles, and reports to parents. As has been stated several times, the document examination was very much secondary to the interviews in the research process. The main aim of this activity was to provide more contextual information for myself in seeking to understand points raised in the interviews and for triangulation of data. Where the documents contribute to a fuller understanding of assessment procedures in the school, this information has been included. The ERO reports for each school were examined in some detail. In the case of Tasman, previous ERO reports were examined in more detail to provide an understanding of the assessment history of the school; this was lacking in the interviews because the three interview participants had only been at the school a short time.

5.12 Data Analysis

The interviews were transcribed and coded using the coding process as outlined in Merriman (1998). Units of information (Lincoln and Guba, 1985) were identified. Using Lincoln and Guba's definition a unit of information must be heuristic; by that they meant the unit should reveal information relevant to the study. The unit should be the smallest piece of information that stands by itself. It should be able to be interpreted without any additional information other than some knowledge of the context. Not all units were classed into categories. The questions on where and how the participant developed their knowledge of assessment were not coded into units. This information provided some interesting background, but after critical analysis was not considered relevant enough to the research questions to be subject to further

analysis. While the process implies a linear progression, the final analysis, for each case study and the integration chapter, was a result of constantly returning to the data for further clarification and both the units and the categories changed over time.

For each case study school, an initial interview was transcribed and units of information identified. The participants' comments on that element of the interview were then coded with that unit. The units were given a name that most adequately described that information, for example: *Assessment for teaching versus assessment for reporting to BOT, ERO and the Ministry (Kate, Endeavour)*. The remainder of the interviews were then coded into those units. Where subsequent interviews in that case school revealed other units, they were added to the original list. The units were then placed in categories using a number of criteria which included frequency and strength of response from all participants, relevance to the research questions, relevance to a particular assessment tool such as STAR, Numeracy Project, and PAT, and illustration of a perspective not shared by other participants.

5.13 Follow-up Activities

Postscript interviews

As stated previously, the interviews of the participants in the case study schools took place during the period November 2006 to May 2007. In May 2009, the three principals of the case study schools were re-interviewed to ascertain subsequent changes in assessment procedures at their schools and the rationale behind such changes. The postscript interview also sought to check the principal's agreement with the summary of assessment themes derived from the original research at their school.

Focus Groups

The focus group interviews with experienced principals formed a second follow-up activity. As stated above, the focus groups were invited to respond to conclusions and explanatory structures developed from the case studies and literature review. Three focus group interviews were carried out with experienced Wellington principals during May 2009. In total 15 principals participated in the focus groups. The process for the focus group interviews is documented in Appendix F.

5.14 Challenges

The permission of each school to participate in this study was granted through the BOT with support from the principal. Once this permission was obtained, participants from within schools (e.g. teachers) were invited to volunteer through an open invitation to all. This proved problematic as no staff responded to this original invitation. This was taken to be an indication of the busy life of teachers rather than reluctance to engage with the research. In response to this, the principal provided the names of staff who he or she thought would be an appropriate participant. These people were invited to participate through a direct approach, but the voluntary nature of the participation was still explained to them. Most agreed to take part.

As stated elsewhere Tasman School presented some significant challenges. All but one member of the teaching staff had been appointed within the last three years. This made it difficult to get a longer term perspective on the rationale for assessment decisions. While there was one staff member who had been at the school for longer than three years, the teacher declined to be interviewed. It was, therefore, decided to only interview three staff members at Tasman, all being members of the management team.

5.15 Trustworthiness, Validity, Reliability and Transferability

Merriman (1998) states “the applied nature of educational inquiry makes it imperative that researchers and others be able to trust the results of the research - to feel confident that the study is reliable and valid” (p. 164). This concept of trustworthiness (Patton, 2002; Robson, 2002) concerns concepts of validity and reliability that were traditionally associated with quantitative research. When applied to qualitative research, validity and reliability need to be reconceptualised, or more precisely, operationalised in a different way so that they can be meaningfully applied to the research process and data. The following sections deal with this for the present research.

5.15.1 Perspective of the Researcher

The first element in the discussion of the trustworthiness of this research includes a transparent statement on the perspective of the researcher. This was provided in

Chapter One and will not be repeated here, but it is important that the perspective of the researcher be understood so that readers can evaluate the research process and findings against this background.

5.15.2 Validity in Research

Validity in research focuses on the fitness for purpose of the research design for yielding data that answer the research questions being examined (Hall, 2007). However there are many factors to consider when establishing the validity of a research study. Merriman (1998) discusses at length the concept of validity in qualitative research. She highlights the fact that any research in this field is mediated through two processes of reality. Firstly, we have the perceptions of the participants being studied. We are recording the participants' perceptions of reality; that is, what they report happened in an event or a process in which they were involved. We have little or no way of checking whether their perceptions are correct. Secondly, these reports are mediated through the researcher's own analysis and conclusion from the participants' original reports.

In responding to the quest for validity in this situation, Merriman (1998) summarises Taylor and Bogdan (1984). "The qualitative researcher is interested in perspectives rather than truth per se, and it is the researcher's obligation to present a more or less honest rendering of how informants actually view themselves and their experiences" (p. 168). Merriman goes on to state:

Most agree that when reality is viewed in this manner, internal validity is a definite strength of qualitative research. In this type of research it is important to understand the perspective of those involved in the phenomenon of interest, to uncover the complexity of human behaviour in a contextual framework, and to present a holistic interpretation of what is happening.

(p. 168)

Merriman lists six methods for achieving a systematic approach to internal validity:

1. Triangulation
2. Member checks
3. Long term observation
4. Peer Examination

5. Participatory modes of research
6. Researcher biases

The validity of this research is discussed in terms of Merriman's six methods.

5.15.3 Triangulation

Triangulation has been defined by Burns (2000) as "the use of two or more methods of data collection in the study of some aspect of human behaviour" (p. 419). Burns goes on to state:

Triangulation contributes to verification and validation of qualitative analysis by:

- checking out the consistency of findings generated by different data collection methods and
- checking out the consistency of different data sources within the same method. (p. 419)

Hall (2007) states that in education enquiry triangulation can be seen as taking place in two main ways:

Corroborative triangulation and coherence triangulation. The first focuses on gathering data from different sources or techniques in order to see if they corroborate each other; all data focuses on the same object or aspect of an object. The second is less about corroboration and more about picture building. The focus is on using different sources and techniques to build a coherent picture of what is going on. (p. 3)

This research contains a number of elements that attempt to provide a sense of validity to the research process and any conclusions that emerge. These elements include those processes that seek to corroborate evidence and others that seek to use a wide range of information sources in order "to build a coherent picture of what is happening." (Hall, 2007, p. 3)

The fact that three schools were used in the investigation and three to four teachers interviewed in each school contributed to building a coherent picture. While it could be argued that more extensive interviewing may have contributed even further to the validity of the process, such issues had to be balanced with the need for the research to be manageable.

The following two paragraphs (and Tables 5.1 and 5.2) provide illustrations of the process of corroboration used in this research. Although the examples pre-empt the findings of the research, this should be seen at this point simply as evidence of the triangulation undertaken both within and across schools.

Comparison of issues arising from all participants in the same school provided opportunities to corroborate the subsequent conclusions drawn from the interview process. For example, at Endeavour School all four participants stated that the writing assessment procedures adopted following the school's literacy contract were time-consuming, but have been of great value in developing all teachers' understanding of formative assessment. Table 5.1 illustrated other common findings from the interviews *within* each school.

Table 5.1: Common elements in data from within individual case study schools

School	Common elements emerging from all participants at the school
Endeavour	The importance of asTTle writing professional development in developing the school's understanding and practices in formative assessment in other areas of the curriculum
Tasman	The effect of the negative ERO report had on contributing to changes in assessment practices.
Kupe	The importance that the portfolio system played in the school's assessment practices.

In contrast, Table 5.2 illustrates some of the elements common to all three schools. The examples illustrate that although each school was unique in its approach to assessment, common elements existed in their development of school-wide strategies. The strength of the research findings is enhanced through such analyses.

Table 5.2: Common elements in data across all three schools

1. The introduction of the concept of formative assessment over the previous two to four years
 2. The important role of professional development in changing assessment practices
 3. The challenge of carrying out assessment for formative purposes while also providing summative data for principals, BOTs and ERO
-

This research used both document examination and semi-structured interviews to build a coherent picture of assessment in the three schools. For Tasman School, where the impact of the previous ERO report had a major influence on changes to assessment practices, it was important to examine the school's five previous ERO reports to gain a comprehensive picture of assessment practices in that school.

5.15.4 Member Checks

All participants were given a copy of the transcription of their interview and asked to confirm their responses or to change any part of the record of the interview. Furthermore in the postscript interview the principal was given a summary of the major conclusions regarding assessment at their school and invited to comment on their perception of the validity of the findings.

Long Term Observation

The manageability of the research itself prevented long term observation and interviews for all eleven participants in this study. However, there was up to six months gap between principal interviews and that of scale A teachers and deputy/assistant principals in the schools. This at least allowed the researcher the opportunity to observe each school on two occasions that were significantly far apart to achieve some perspective on the stability of the phenomenon over time. For example, if assessment processes or assessment tools had changed then these would have been identified and considered in the light of the information obtained from interviews.

Postscript interviews were carried out with the three principals in April 2009. These interviews discussed subsequent changes in assessment processes as well as allowing

the principal to comment on conclusions the researcher had made regarding assessment at their school.

Focus Groups

In 2009 the major conclusions of the research were examined by three focus groups of principals. The comments of these 15 experienced principals led to some modification of the research conclusions.

Peer review

Peer review was limited to comments made by doctoral supervisors.

5.15.5 Participatory Models of Research

Initial research findings for each case study school were shared with the teachers and principals; where pertinent comments were made, this has been included in the final discussion.

5.15.6 Researcher Bias

The concepts associated with researcher bias have been covered in Chapter One (see section 1.7).

5.15.7 Reliability

Whereas validity focuses on fitness for purpose of the research design, reliability is concerned with the extent to which the results of the study provide an accurate picture of the behaviour, beliefs, attitudes or performance of participants in relation to whatever is the focus of the research (Hall, 2007). According to Hall, in traditional normative/positivistic research designs the notion of reliability is operationalised through the concept of “consistency” or “repeatability”. If the results of a study are replicated the inference is then drawn from this consistency that the results can be relied on as an “accurate” representation of whatever was measured or tested. However, qualitative research with its in-depth study of highly contextualised environments prevents the application of the traditional notion of “reliability as consistency”. As the research largely, but not exclusively, deals with people’s perception of reality in their specific context, it is not reasonable to expect findings to

be replicated if the same approach is used in a different context with different people. Similarly, studies are often conducted on phenomena or events that happen only once because the same combination of people or circumstances does not occur again (e.g., research may be undertaken on unruly crowd behaviour to identify steps that could prevent such behaviour again). With qualitative research, the concept of reliability is operationalised differently. The key notion is that of “reliability as dependability”. In this sense research is considered reliable if the processes being used have been systematically applied and represent actions that can be “depended” on for their soundness. Dependability is very much about attention to detail: the research is systematic; the views of participants are accurately recorded and interpreted; member checks are conducted; triangulation is appropriately undertaken to confirm data and interpretations; and a “rich” and “thick” description of the process and research context is provided so that other researchers (or even readers) are able to “audit” what has been undertaken.

For this research, the following procedures provide evidence of reliability as conceptualised within the notion of dependability:

- Rich or thick descriptions are provided;
- Triangulation was undertaken in the manner described earlier;
- Member checks were carried out;
- Peer review was undertaken by the researcher’s supervisors.

5.15.8 Transferability

Traditional scientific method is concerned with discovering laws in a particular context that can be applied across a wide range of behaviour; predictability is the goal. This ability to generalize and predict is based on controlled experimentation and the use of representative samples. Clearly, the nature of qualitative research does not lend itself to controlled experimentation and the use of representative sampling. For these reasons, the researcher is not in a position to generalise the findings from his or her study to a wider context. Instead, it is the responsibility of other researchers, or readers of the research, to decide whether the results transfer to their own context. The notion of “transferability” therefore replaces the notion of “generalisation”.

Merriam (1998) lists three ways that transferability, which she labels as “reader or user generalizability” (p. 177) can be enhanced. Researchers should provide detailed descriptions of the research process in order that those interested in applying these findings to their own context have a comprehensive base of information on which to make a judgment. She also recommends that researchers describe fully the context of the study so readers can make an informed judgment on how closely their own context aligns with that of the research. Merriam’s final point is that researchers should undertake, where appropriate, multi-case studies in order to provide a greater range of contexts. All three points have been addressed in this research.

5.16 Ethical Issues

The consideration of ethical issues during the design and conduct of the research are a vital element of any research undertaken that involves people or animals. This study was approved by the Ethics Committee of Victoria University. As Merriam (1998) states: “Professional codes and federal regulations deal with issues common to all social science research - the protection of subjects from harm, the right to privacy, and the notion of informed consent” (p. 183).

The interview participants were not put at any risk by the process. However, the issue of privacy did present a number of challenges. All schools were given a pseudonym to disguise their identity. However, Wellington has only 140 primary schools and while every endeavour was made to disguise identity, it may be possible that local teachers and principals could identify the school. Considerable thought went into the choice of names. It was finally decided that the paths schools had taken in assessment were very similar to the history of voyages of discovery to Aotearoa/New Zealand. Schools had only a vague idea of where they were heading and few maps to guide them. They were often blown off course by adverse winds and currents. For this reason the schools were named after explorers to this country or their ships: Kupe, Tasman and Endeavour. The participants were all given pseudonyms.

While it would have been desirable to keep the identity of the scale A teachers being interviewed hidden from the principal and other senior management, the reality of school life prevented this from being practical or possible. While these issues do pose somewhat of a problem for both the schools and individuals involved in the study, it is my understanding that the schools and people were aware of these matters when they

consented to participate in this research. All tapes, transcripts and documents were kept in a secure location at my house and will be destroyed two years after the completion of the study.

Informed Consent

The principals of all the schools were given full explanations of the nature of the research in a telephone call. Once they agreed to be involved in the study, formal written approval was sought from each school's Board of Trustees.

Before each interview the participants were given a written explanation of the research and the ethical guidelines associated with the study. Written consent was obtained from individual participants (Appendices C and D, pp. 167-168). One potential participant declined to be involved in the research after the study was explained.

Deception

All participants were given a copy of the transcript of the interview and were given an opportunity to change or delete any comments. A summary of the research for each case study school was also given to all participants at that school with an opportunity to comment.

5.17 Summary of Methodology

This chapter outlined the research methodology used in this study. The study was firmly located within an interpretative paradigm and involved the collection of mainly, but not exclusively, qualitative data. Quantitative data involved recording of the assessment procedures administered and the frequency of their use. The qualitative data focused on attempting to understand teachers' and principals' perceptions of the manageability and utility of the assessment systems and tools used in their schools. The research was conducted in three schools which represented the bounded systems under investigation. The study used research methods and data collection common to many forms of qualitative research; the dominant method of data collection was the use of semi-structured interviews with school staff, with other data being obtained through document analysis and subsequent focus group discussions. The chapter also outlined the process by which the study attempts to provide a trustworthy description of assessment issues in New Zealand schools.

CHAPTER SIX: CASE STUDY SCHOOLS

6.1 Introduction to Chapter Six

This chapter begins with a brief outline of the research processes in the three case study schools. This is followed by a description of how the information gathered has been organised to provide consistency in analysing each school's assessment processes. Each school is then discussed separately beginning with a statement and table that presents some contextual information to provide a fuller understanding of the school's situation.

Research Questions

For the convenience of the reader, research questions two and three are restated.

Two

How did three New Zealand primary schools establish assessment and reporting systems during the period 2003-2007 that were manageable and maintained a focus on learning?

Three

To what extent do the staff of three case study schools find the most recent changes in assessment practice/tools manageable and provide useful information?

6.1.1 Research Process

As stated in the previous chapter the major element of data gathering was largely based on semi-structured interviews that took place between late 2006 and the first quarter of 2007. In addition to the four interviews, assessment documentation in the school was examined. This included policies, reports to the BOT, student portfolios, and report forms. The schools' past ERO Reports were also used as part of the data gathering process. However, as noted in the methodology chapter, the most dominant element of the data was the interviews. In 2009 the principals were interviewed again. They were asked to comment on their understanding of how well the conclusions about the major themes of assessment in their school matched with their experience.

They were also asked to discuss the rationale and process relating to changes in assessment processes at the school over the last two years

Four teachers were interviewed at Endeavour school: the principal, deputy principal and two senior teachers who had recently been given management responsibility for assessment at the school. At Kupe school four people were also interviewed: the principal, a syndicate leader and two scale A teachers. Selecting interview participants at Tasman proved to be more problematic. Apart from one teacher all staff at Tasman had joined the school within three years. The only teacher who had been working at the school for longer than three years declined an invitation to be interviewed. It was therefore decided to interview only the three management staff at Tasman. It was for this reason that there was a deeper analysis of Tasman's ERO reports than with the other two schools. It was felt that more in-depth analysis of ERO reports at Tasman would give a more comprehensive background to discuss assessment at this school.

6.1.2 Structure of the Case Study Report

The interview schedule is included in Appendices A and B. The interviews in each school were transcribed and coded using an open coding process as outlined by Strauss and Corbin (1998). This is covered more fully in Chapter Five. This process included close analysis of the individual transcripts to identify key elements in the participants thinking on those areas of assessment most related to the research questions. The data from the individual interviews were then combined to identify common elements that were shared by the participants at that school. Finally, the coded data were analysed to develop key frameworks for that school. While the frameworks were similar for the three case study schools, the data analysis did produce some differences in the key elements of each school's assessment processes. The differences reflect the fact that no two schools are the same in respect of context, history, staffing, students, systems and choice of assessment tools. Tasman School in particular has a history that is significantly different from, and more challenging than, the other two case study schools.

6.2 Case Study One: Endeavour School

6.2.1 Description of Endeavour School

Endeavour is a decile two state school with 330 students, 80% of whom are Maori or Pasifika. Richard the principal has been at the school over ten years and Grace the deputy principal has been at the school close to twenty years. The school has a stable staff and has had a succession of positive ERO reports. Endeavour is a full primary school (years 1 to 8).

At the time of the research there were 16 teachers at the school including a non-teaching principal. The school was organised into four syndicates. This was made up of a four class junior school of two new entrants' class, one year one class and one year one-two class. In the year two-three syndicate there were four classes. This included three year two-three classes and one year three-four class. The third syndicate was comprised of two year four-five classes and two year five-six classes. In the Senior School there were three year seven-eight classes.

The major decisions regarding assessment were made by the principal and the two designated assessment leaders. These three consulted with other staff members when appropriate.

The semi-structured interviews conducted at Endeavour were with four staff members: the principal, deputy principal and two teachers. These teachers had been allocated the role of assessment leaders in the school in addition to their classroom teaching roles. It was decided to interview the assessment leaders rather than two scale A teachers because they had the most experience with assessment issues at the school. Richard, the principal, was interviewed in November 2006 and the three teachers in February 2007.

Table 6.1 provides a summary description of Endeavour School and Table 6.2 provides a description of the assessment procedures at February 2007.

Table 6.1: Description of Endeavour School

Roll	333
Decile	2
Structure	Full primary
Ethnicity	Maori 40%, Samoan 16%, NZ European/Pakeha 11%, Tokelauan 13% Cook is Maori 9% Asian 5% Tongan 3% Indian 1% Other 2%
Principal	Richard
Deputy principal	Grace
Assessment leader	Susan
Assessment leader	Mary

Table 6.2: Assessment and Reporting Practices at Endeavour School as at February 2007

ATOL Contract	No
Literacy Contract	2004-2006
PAT Tests	No
STAR	2004-2006 once a year, 2007 twice
asTTle Writing	from 2004
asTTle Reading	2006 – some use senior school, 2007 formal use year 4 and above
asTTle Maths	No
Numeracy Project	In classroom not analysed and used school-wide
The Exemplars	Writing used formally in Junior School as part of school wide process Some informal use of other Exemplars
Spelling	Daniels and Diack
Basic Facts	School developed (post 1996)
Running Record	Yes
ARB	Informal use
SEA	From 1998
Six Year Net	Yes
NEMP	No
Cumulative Cards	Yes
Portfolios	Yes
Quality Workbooks	Yes [a self assessment book that is based on the philosophy of the Quality Schools movement]
Student attitude surveys	Yes
Written Report	One mid year; year 8 December as well
Parent Interviews	Yes
Assessment software	<i>Classroom Manager</i>

6.2.2 Assessment Prior to and in the Early Years of the Reforms

Both Grace, the deputy principal and Richard, the principal recalled that there was very little assessment being carried out at Endeavour and other schools both before the reforms and during the early 1990s. In Grace's words:

There wasn't much to tell you the truth. We did Running Records. I don't even think we had formal maths assessment. We didn't have formal writing. There was nothing school-wide.

Richard described his time at another school during this period and his arrival in 1996 at Endeavour.

I can remember P and A Registers before Tomorrows Schools came in. That's all the formal stuff there was, not a lot of anything else. Well the schools I was at never looked at school-wide trends and things like that. It was only within your classroom and your syndicate where you looked at what kids were doing.

When I came here in 1996 there was no school-wide assessment, so I started to put those sorts of things in and strengthen practices within the school so they were much more consistent around the school. Rather than individual teachers doing different things which was happening before I came. A lot of schools were like that, I am sure.

6.2.3 Endeavour Assessment Practices, February 2007

While Grace and Richard reported that there was very little assessment activity being undertaken at Endeavour School in the early and mid 1990s, by February 2007 the school was undertaking a large amount of assessment developed in various ways over the preceding 10 years.

At the outset it needs to be stated that the use of asTTle Writing and the Literacy Contract undertaken in 2004/2005 dominated thinking and activity in assessment at Endeavour. The professional development in the teaching and assessment of writing was in the process of being embedded in the school. The four participants all stated that this development had provided the school with its current understanding of formative assessment. There was a consensus among the four staff interviewed that although assessment in this area was very time consuming it provided very valuable information.

Mary, one of the assessment leaders describes asTTle writing:

Yes, it's been a huge process for us with a lot involved. It has been a great process actually because along with the professional development that we did it meant that our teaching was a lot better and that's not a static thing either. It's a very useful tool and it shows where those children are and you can quite clearly see those areas that you want to look at.

These views were supported by Richard the principal:

We have been on the writing journey for two years with intensive professional development. We are trying to embed a lot of the things from that. We have used AsTTle very comprehensively and our expertise in that area has grown immeasurably including moderating kids writing and that has sort of spilled over to the reading experiment for this year.

6.2.4 Assessment Goals

To understand the establishment of assessment and reporting processes at Endeavour, it is helpful to begin by examining what the school attempted to achieve in this area of their activity. While not formally documented in the school's policies and procedures, the analysis of the responses of the four interviewees identified eight goals that the school was trying to achieve in its assessment reporting procedures:

1. A manageable system
2. The use of assessment tools with clearly established purposes that meet the school's needs
3. Whole school assessment procedures
4. Consistent assessment practices
5. A system with a strong commitment to formative assessment and the use of analysed data to move children forward in their learning
6. A system that met the needs of its specific school community, in this case, a decile 2 school
7. Procedures that met the legislative requirements of the Ministry, the school's obligations to the BOT and the desire for a positive ERO report
8. A desire to keep changes in assessment procedures in perspective and not let assessment dominate the culture of the school

A number of these goals such as consistency and a school-wide system were closely related. While some of the goals such as formative assessment left a strong footprint

in the interviews, others such as the desire not to let assessment dominate the school culture were not as explicitly stated. Each of these eight themes is briefly discussed.

A Manageable System

In talking about manageability, Grace stated the school's goals succinctly when she said: "All the senior staff wants, is to do the assessment in the easiest, most practical way, in a way that is not time consuming."

The Use of Assessment Tools with Clearly Established Purposes

A major element in making assessment more manageable is only to undertake assessment that has a clear purpose. This was a strong theme in the interviews of the two assessment leaders Mary and Susan. It is clear that the assessment leaders saw reviewing the purposes of various assessment procedures as one of their major roles in the school. As Susan stated "We can do this assessment but what is the purpose? Why are we doing it and what are we going to do with the information we find out?"

Mary supports this view when she states:

The whole issue, and it's not static, you have to question why you are assessing - what is it you are trying to find out? Are you trying to find out the best way you can? Once you have found out what you are trying to find out what are you going to do with that?

School-Wide Procedures

The two longest serving members of staff, Richard and Grace, both made the point that in the early and mid nineties a whole-school approach to assessment was lacking. A number of the changes in the assessment procedures since then have been aimed at adopting a whole-school approach to assessment. In response to being asked to summarise what he thought were the most important things schools need to have in place in terms of assessment, Richard responded: "Number one is a school-wide picture so you can see trends on a school-wide basis." This approach is most evident in the writing assessment, Quality Workbooks and mathematics basic facts. This goal does present a number of challenges which will be discussed later.

Consistent Assessment Practice

Closely associated to a school-wide approach is the desire to have consistent practice across the school. Grace described how this movement toward more consistent assessment practice had developed in the last few years, especially in response to ERO recommendations: “It was just more getting consistency with our assessment. Consistency was the key and there were areas that we weren’t assessing school-wide that should have been.”

The issue of consistent practice across the school includes the interpretations of results as illustrated by Mary in the assessment of writing. The school uses asTTle Writing for assessing students in years 4 to 8 and the Exemplars for children from new entrants to year 3. There is a challenge to get consistent scoring when using two tools.

There is a bit of a conflict, well not really a conflict, but making sure the Exemplars and asTTle match. So if you are talking about a level two from the Exemplars well then making sure that we are being consistent about two different types of marking criteria. Looking at the criteria for the Exemplars and looking at what asTTle says because they have different criteria.

Susan describes consistent practice as one of the challenges the school faces:

I have found some of the challenge is getting consensus from teachers and getting uniformity. As well as making sure people are doing exactly what they should be doing.

In another section of her interview Susan identifies numeracy assessment as an area where consistency is an issue: “I have just had a discussion with Richard about that. The Numeracy Project assessment I suspect, in fact I know, is not being used in the school as it should be as it is set down in our guidelines.”

Commitment to Formative Assessment

The language of formative assessment dominated all four of the interviews. The Literacy Contract and the introduction of asTTle Writing was the vehicle for the emphasis on this goal.

As Richard states: “We haven’t done ATOL, however a lot of what we have done in the Writing Contract has been ATOL type stuff. Learning intentions, success criteria that sort of thing etc. Doing that in writing has taken that out and spread it through the

curriculum.” This is reinforced later in the interview when Richards states “the Writing Contract has made the teachers more aware of formative assessment and has made them formalise it.”

Meeting the Needs of the Community

It is clear that there is a desire from the school to meet the needs of its specific community. Endeavour is a low decile school with a high percentage of Maori and Pasifika children. The school’s response is reflected in the school’s long-standing rejection of PAT tests which they deemed to be culturally biased. It is also evident in the efforts that the school puts in to get parents to attend interviews. As Richard notes:

There is a written report that everyone gets mid year followed by interviews which we tend to get 100% turnout which we see as key for goal setting and making plans for the rest of the year. We follow parents up, we just about drag them down here, and we pick them up sometimes to get them down here.

Meeting Legislative Requirements

Like all New Zealand schools Endeavour has obligations to meet legislative requirements in terms of planning and reporting to the Ministry of Education, reporting on student achievement to the Board of Trustees, and being subject to the audit process of the Education Review Office. As Mary states about ERO – “You are always trying to have what they are expecting.”

Teachers spoke a number of times about the amount of assessment data that needed to be gathered by the principal and reported to the Board. Richard noted how this reporting had increased over time:

It has changed because the BOT are now getting a lot more curriculum information. They are spending a lot more of their meeting time on curriculum things which is a good thing. They get all our school-wide achievement data and there is a lot of it. We talk about it and they do ask questions about it.

Procedures have been put in place to implement the planning and reporting requirements (MoE, 2001a) which were developed to include more staff ownership of the target setting process. As Susan describes the present process:

We have just set our targets and they came up for discussion at the initial meeting the other day. We have gone out to the community, we are not entirely happy with the targets that have been set and they are going to be up for discussion before they are in place.

Mary's description of the previous year's process illustrates some of the challenges around the new requirements:

Richard had to go on this course for a couple of weeks and we must have already discussed the targets for the year. We must have had some sort of rough idea about what we were going to do. I honestly can't remember. So we came in on Monday and there were the targets. We were supposed to talk about them in our syndicates, but we had curriculum team meetings anyway and then senior staff meeting in the morning but we barely had time to breathe you know and all of sudden you have these targets. And to be fair at this particular time, I think we all felt by looking at them, they all looked incredibly high. I am not saying you don't aim high. We just felt that they were very, very, very high. Because it affects everybody, because here am I as a senior teacher saying to these guys, right we have a set of year fives, this is where they are and this is where they have to get to by the end of the year. Which is great because it is good for your professional development and it improves your classroom practice. But also you also have to make sure what you are asking for from your teachers, and all your children, is actually realistic, otherwise you frighten the hell out of everybody and everybody thinks they have done a crap job.

While the school is motivated to meet Ministry of Education requirements and obtain positive ERO reviews they are relatively relaxed about the influence of outside agencies. A series of positive ERO reports and feedback from the Ministry has acted to confirm the school's practice rather than implement a great deal of change in practice.

Mary described past ERO visits: "For the one before that; if I remember correctly, they wondered why we were collecting so much information and I think that had an effect on the portfolios and the Quality Workbook that we now have."

When questioned on any influence ERO had on school assessment practice Susan stated: "They did not in the last visit but in the visit before when they came in and said we could do some things better and change the way we do things."

Richard described the series of ERO reviews in the school in a positive manner when asked if ERO had influenced school practice in assessment:

They have probably tried (laughter) no not really because they have always been really happy with what we have been doing, only superficial, the review we have just had they looked at all the things about our assessment and suggested some tweaking of things but in all we got an area of recommendation for improvement which was focused on writing which brought up a few subheadings and all those things we told them we are doing, so there is no issue with what we are doing at this school.

We have gone our own way in making decisions. Some of the early reviews. The first one I had only been here five minutes, in my second term. The second one was where they were most critical but we stood up to them we had reasons for being where we were at that stage, some were personal some were other events that put us back and the last two were very wonderful reviews basically.

Keeping Assessment in Perspective

In his leadership role Richard demonstrates a relatively relaxed attitude to change.

On the chance of the school doing an ATOL Contract he states:

I don't think we need to. It would be a bit over the top. A letter came about ATOL the other day. It went straight in the bin. We have covered it in other ways. I don't think we need to do ATOL formally. You can't bore teachers to death.

In describing assessment changes in the school since 1996 he stated: "It was more of an evolving process rather than a huge deliberate focus on assessment. There was no whole year focusing just on assessment."

6.2.5 The Implementation of Assessment

From reporting very little assessment activity in the early nineties, by 2007 the school was operating a comprehensive and complex system as shown in Table 6.2. Analysis of interview data gives greater understanding of the implementation and modification of assessment activity. The use of student portfolios, a larger number of Running Records, school-wide use of basic facts tests, and cumulative cards were developed under the leadership of Richard after he was appointed in 1996 and discovered a dearth of assessment in the school. As stated earlier Richard describes this process as an evolving one developed from reflection among staff and informal contact with other schools.

In the last seven years there were a number of assessment activities that were experimented with or informally trialled by individual teachers or syndicates but did not develop into school-wide processes. This involved use of the Exemplars in areas other than writing, the Assessment Resource Banks and to a lesser extent the NEMP Tools. Teachers who experimented with these tools can be seen in some ways as explorers, searching out new land and reporting back to the rest of the group where

something of value had been discovered. Mary illustrates this in her comment on the use of NEMP:

We have had a look at them – I know one block has used them, something from the writing, I have had a look and they have been part of the project. They are quite interesting to look at. You get a feel about your school and other schools.

When Susan was asked about how widespread the use of the Exemplars was, apart from their use in the assessment of writing, her response illustrates the informal nature of the use of these tools: “more usually, just me in my classroom. I don’t know how widespread other people use them.”

Richard makes a similar observation about the use of the Assessment Resource Banks: “We don’t use them as a school but some of teachers have used them for some of the assessments they have done within the class. Some have looked at them and used them.”

This informal use of an assessment tool can lead to it being developed further in the school’s more official assessment activity. This was the case with the introduction of asTTle Reading. Susan described it as being *slipped in* by a teacher rather than going through what have become the school’s more formal processes of implementing new assessment procedures.

asTTle Reading is only something we introduced last year and it was kind of brought in by another teacher who slipped it in and realized too late that they had gone too far and had not consulted with the two most crucial people about how it was important that we knew about it before it went to the staff but by then it had been discussed by too many people who thought it was going to be really good.

In contrast to the introduction of earlier assessment procedures such as the profiles and basic facts tests, which were introduced through an evolving process and were largely developed by Endeavour themselves with some informal contact with other schools, asTTle Writing, Numeracy Project Assessment and the Quality Workbooks came from external sources and were introduced in a more direct process. asTTle Writing has become the most dominant assessment tool at Endeavour in terms of how much time is devoted to carrying out the process and how frequently it was referred to by the four staff interviewed. Although teachers all describe the processes involved with asTTle Writing as being very time consuming, the assessment is very largely described positively. As Mary states: “asTTle Writing has been particularly valuable.

As part of it has helped teachers with their professional development and become better practitioners.” As stated earlier, Richard noted that professional development in writing had contributed to using the language of formative assessment in all areas of the curriculum. While asTTle Writing has, in the views of the interview participants, been successfully implemented and has become embedded as a central assessment procedure at Endeavour, the same could not be said of the assessment activities associated with the Numeracy Project, which was also a Ministry funded contract. Its introduction has been less positive and it has not yet become embedded in the school’s practice.

This is illustrated in a number of ways. For example the new entrant teacher has decided to retain the number elements of the School Entry Assessment rather than using the Numeracy Project Assessment. As Susan notes very directly: “The Numeracy Project assessment is not being used in the school as it should be, as it is set down in the guidelines.”

Teachers frequently mention how time consuming this assessment is. However, Richard puts a lot of the problem down to the structure of the professional development contract.

We did the Numeracy Project piecemeal; the Junior staff, then the Middle staff, the Senior Staff and in different years and it was never like the whole school, everyone at the same time. When it is whole school I sit in on the meetings and learn along with the teachers. It was off-site, a lot of it. I didn’t get to have much hands-on stuff and it’s been very piecemeal.

Richard also struggled to see how he could use Numeracy Project Assessment for school-wide reporting:

I have seen the results the teachers have given me - the sheets they have done. And so I think to myself how can I use this as a school-wide thing but I can’t get my head around that and work out what to do.

The Quality Workbook came out of a group of local schools adopting the Quality Schools Philosophy which operates in several regions in New Zealand. Richard credits it with introducing more self assessment into the school.

Review

The school staff is constantly reviewing its assessment procedures. This activity has probably become more obvious since the appointment of the two assessment leaders. In this role they are influenced by issues of manageability, the purposes of assessment and the use of school management software. Susan describes how the assessment leaders have worked together in reviewing assessment:

Mary and I have identified changes in our current assessment procedures that we think are necessary. The main one is the cumulative cards and how we use those, what we use those for and we are also implementing *Classroom Manager*. Why have another piece of paper that effectively nobody looks at and then goes back to. What is the purpose and why are we using it, how is it going to inform what is going to happen?

There is also a desire to review procedures in order to establish more consistent practice. This is evident in Mary's concern over the use of the Numeracy Project Assessment. Current developments in education are also being flagged for implementation such as the Key Competencies and the assessment of thinking associated with the new curriculum.

For Richard the review of procedures also includes areas where he feels the school lacks sufficient information. While the school assesses mathematics using its own basic facts test he feels they are "a bit light" in this area.

Summary of the Major Themes Regarding Implementation

In summary, this section of the analysis of the interview data indicated six major themes in the implementation of assessment practices at Endeavour School.

The assessment procedures introduced were related to the school's assessment goals.

Assessment procedures developed in the early period (1993-2002) were more informal and school-based.

Assessment activity in more recent times has been more formal (asTTle, STAR, Numeracy Project) and influenced by professional development contracts supported by the Ministry of Education. Since 2000 the Ministry has become more active in developing and promoting assessment tools.

The more successfully implemented assessment procedures become quickly embedded into school practice. Other new assessment activities spend a longer period

in the transition phase before procedures become consistent and there is a sense of staff ownership. An assessment practice is more likely to become embedded practice if it is associated with high quality and well structured professional development.

Assessment activity has always been under review. The pace of that review has increased with the appointment of the two new assessment leaders.

Assessment software is expected to have an increasingly important role in assessment practices within the school.

6.2.6 Challenges in Implementation

The research question focuses on the challenges associated with the establishment of assessment systems in the case study schools. This will be discussed with particular focus on the implementation of the most recent developments in assessment.

In considering the manageability of assessment in schools the most obvious issue is the time required to undertake assessment. While this aspect did feature strongly in the Endeavour Case Study and will be explored shortly, the interviews revealed another significant manageability issue related to the mismatch between some assessment tools and the way the schools are structured.

The Endeavour 2004/2005 Literacy Contract led it to develop a whole school commitment to the assessment of writing with asTTle being the major assessment tool. The whole school approach involved undertaking writing assessment at the same time, using the same writing topic and a moderation team drawn from all syndicates in the school. While this approach has many positive aspects, the fact that the asTTle Writing test is designed to only assess children from year 4 and above does present a challenge to managing the assessment of writing. Classes below year 4 use the Exemplars for writing assessment. This means that those composite classes with a mixture of years 3 and 4 present a major challenge at assessment time. While both tools are based on levels of the New Zealand Curriculum, asTTle and The Exemplars have different performance indicators in their marking schedule. As Mary states:

There is a bit of a conflict, well not really a conflict, but it is an issue making sure the Exemplars and asTTle match. So if you are talking about a level two from The Exemplars we then mark them trying to make sure that we are being

consistent about two different types of marking criteria one from The Exemplars and the other from asTTle.

The commitment to the whole school writing assessment regime also presents a challenge in its requirement to use the same writing topic for formal assessment throughout the school. As the teacher of the youngest children in the school, Grace sees this as an issue.

There is vigorous debate every year on the writing assessment (laughter). I think that will continue, mainly to do with once we have chosen the genre, it's the topic. That's always been a headache because I am pushing my barrow to get something that suits my new entrants. It has to be something they know about and it has also to be something that suits the whole school so that's always an issue.

The vast majority of New Zealand primary schools teach the non-core curriculum areas of science, social studies, health and technology in blocks of 5 to 8 weeks often under the collective name of *Topic*. Such an approach presents a major challenge to the utility of assessment. At Endeavour each syndicate develops its own assessment procedures for these areas. The results are collated and given to the principal who uses them to report to the BOT. Other than that, the assessment appears to have little value because the six week topic approach means that by the time the children are assessed and the results analysed the programme has moved on to the next topic. Mary outlined this issue forcefully:

When Susan and I went on our last course and you got a choice on what workshops you went to. There was one on assessment in Social Studies and actually I couldn't help myself I sat through it. It was all very interesting and there was lots of relevant things said but in the end I said sorry but I can't help this. The whole thing was that you have assessed them, so now you have some information about their levels and their next steps but I said quite frankly for us there is no next step. We have done it and you are not going to re-teach it, you are not going to revisit you have done it for this year and it's gone. That's the reality. I got the feeling I should not have said it but that is the reality.

6.2.7 The Manageability and Utility of Individual Assessment Tools

As stated above the most obvious manageability issue in a school undertaking assessment procedures is the amount of time required to carry out the testing and

analyse and record the results. The most appropriate manner in which to examine this issue is to discuss the individual assessment tools.

asTTle Writing and the Writing Exemplars

asTTle Writing (year 4 to 8) and The Writing Exemplars (new entrants to year 3) were introduced into the school in 2004 through a two year Literacy Professional development. asTTle is featured most strongly in the interviews of the staff members. Its introduction into the school illustrates the relationship between teachers' perceptions of the utility and manageability of assessment tools. The four teachers interviewed all stated that their school assessment procedures in Writing (asTTle Writing and The Exemplars) were very time consuming. However, the utility of information that the school's writing assessment provides in terms of giving feedback for children, assisting the quality of teaching and providing summative data for tracking school-wide achievement, was such that the extensive time required was accepted by the staff. In addition to the utility of the assessment data, it was obvious that Endeavour School had a common vision on the teaching and assessing of writing developed from what was reported as a very well led and productive professional development contract.

Numeracy Project

The Numeracy Project was introduced into Endeavour school in a series of one year professional development contracts. The Project has three formats for assessment – what is termed the full number assessment which takes an estimated 25 minutes per child to administer and the abridged version termed a GLOSS. The third assessment technique associated with the Numeracy Project is termed ICAN which is a short paper and pencil assessment whereas the former are based on oral interviews. This quote from Susan captures the concern with the time required for assessment.

I was going to retest all my children using the whole number assessment to see how long it took me before I launched into all the staff to say they should do it. I wanted to know how much time it would take. Because I know what is going to happen. Because there will be a riot we haven't got enough time to do that. I embarked on that last week and I have managed to do two of the whole assessment. So what I have done is looked at it again and I used the Gloss format and I know in my own syndicate two of us have used Gloss in the last two days to find out where the children are at to group them. The third person in our syndicate is going to start doing it.

We did have it originally that every child would have the full assessment twice a year when we were on the Numeracy Project.

That was one of the things that you did as part of the training and after that depending on where you came in depended how you fixed those things. We kind of continued with that because we then had the problem of new teachers coming in all the time and they were having to do that and we kept with that. And I said it is hopeless trying to do that every year as well. The Gloss makes it much easier to find where children are.

As stated earlier in the chapter, Richard has at this stage, not found an approach to use the accumulated data, and the school at the time of the interviews was unclear on its testing regime in this area. Both Richard and Susan have, however, noted mathematics assessment as an area needing review.

There is, therefore, in Richard's view a sharp contrast between assessment with asTTle Writing and assessment associated with The Numeracy Project in terms of the utility of assessment. Both are reported as being very time consuming. However, the assessment process and information associated with the Numeracy Project was not as valued as the Writing assessment and the school lacked a common vision on the use of these assessment procedures.

Assessment of Reading: Overview

In the junior school the assessment of reading is largely unproblematic and based on procedures which have been embedded in New Zealand schools for over 20 years. This includes some early identification based on School Entry Assessment, Running Records and the Six Year Net Observation. Grace does mention some concerns over the consistency of the administration of Running Records but apart from this there are no major concerns regarding utility and manageability.

The assessment of students reading from year 3 and above is more complex as the school begins to make use of two recently developed reading assessments, STAR and asTTle, and some debate about the use of Running Records for assessing children of this age in our schools. While the school's introduction of asTTle Writing was deliberate, well-structured and involved close involvement of the Literacy Contract Facilitators, assessment activity around reading at Endeavour illustrates a more informal and less directed changing of assessment procedures. Richard's description of reading assessment captures this transition period.

We have been doing the STAR reading tests - yes and for the past few years only doing it once a year, a snapshot of achievement but the decision has now been made for next year and I see just this morning some schools are doing it three times a year, or two and a half times a year. Which is something we could look at later on. So we have been doing that for the last four years with STAR. In terms of reading we have also been doing Running Records throughout the school but I have been thinking about a better way; especially with kids who are really good readers. What does an actual running record tell you? Not very much at all. So the STAR is better for them and for this year we have started experimenting with the asTTle reading test without a lot of, without any professional development, so we are doing our own, so we are not experts at setting tests, and we are not experts at analysing information.

STAR

STAR was introduced in 2004 and originally carried out once a year but the school has now decided to administer the test twice a year, hoping that this change will give the school an additional way of measuring progress. The STAR assessment gets a very positive report on manageability and utility. As Mary observed: “We can see how kids are doing and that’s really good. It is easy to administer and easy to mark and stick on *Classroom Manager* and easy to get the results out.”

Running Records

All four teachers interviewed had a slightly different understanding about what were the school’s expectations on the administration of Running Records. The school had decided recently to reduce the administration of Running Records from three times to twice a year. This change arose out of a concern regarding how much time was needed to administer this individual assessment. Of the four respondents, only Richard raised concerns about the utility of this assessment with older children.

asTTle Reading

As stated earlier asTTle Reading had been informally introduced into the senior classes in 2006 and was being used in a more structured manner in 2007. asTTle Reading’s introduction is an example of a teacher acting as an explorer referred to earlier in the chapter. The teacher had experimented with the use of the test in her class, told others and its use became more extensive. One of the assessment leaders referred to it as having been *slipped in* and Richard noted it was being used without any professional development.

An interesting issue in a number of New Zealand schools is comparisons being made between the use of STAR and asTTle Reading. While Richard rated STAR highly for

its manageability, asTTle was perceived positively in terms of utility of the information it provided teachers. As Richard states

In the senior school the asTTle and the STAR, particularly the asTTle for higher achieving children, which we have a number, will provide more detailed concrete information which teachers can actually use to develop programmes to meet the needs of the kids.

Individual Profiles, Cumulative Cards and Quality Workbooks

Endeavour's system of tracking individual children in a number of curriculum areas using Individual Profiles, Cumulative Cards and Quality Workbooks would seem complex and somewhat time consuming. The system appears to duplicate assessment by recording similar information in three different documents. The three teachers, and to a lesser extent the principal, report challenges with these three procedures around the issue of consistency of data collection, manageability and the utility of non-core curriculum assessment that have been documented fully above. Of the three, the two assessment leaders Mary and Susan have identified the cumulative cards as being hard to justify. This issue was also raised by Grace "Each child has three including the cumulative records. Some schools don't use the cumulative cards do they? I think Richard is going to bring up some debate about whether we should continue".

6.2.8 Assessment Software

In 2006, the Ministry of Education mandated that for over the next two years all New Zealand schools had to be using an approved school management software package which included an assessment element. Endeavour has chosen *Classroom Manager*. The school is in the process of implementing the recording of assessment data through this program. The biggest reported benefit so far has been the decrease in time the principal requires to collate summative data. Richard describes how it has decreased his workload:

It is making a huge difference to my workload in the school, school-wide stuff. I used to get handwritten data for each kid and the teacher and I used to look at it and used to have hand drawn the graphs and write all the bits and pieces which for school wide for one assessment took hours and hours work and multiply that by three and I was busy at weekends and what have you. The *Classroom Manager* program has particular idiosyncrasies which you have to learn your way around but once you have mastered that they tend to work pretty well.

Mary, the assessment leader, with responsibility for implementing *Classroom Manager*, is also very positive about the program's utility:

We have tried to make *Classroom Manager* (CM) useful to us as a data gathering tool. Part of the process has been the accumulation of data in the one place, making it easier for Richard to access the data and make his graphs and report back to us. I like it. My role is to look at trends and patterns and anomalies and being able to compare that.

6.2.9 Formative Assessment

The language of formative assessment is dominant in all four interviews with the participants reporting a very positive attitude to this approach. However, Mary does raise manageability concerns over the gap between the reality and ideal in using formative assessment. She responds to suggestions from a recent ERO reviewer about writing formative assessment comments on all children's work in all areas of the curriculum.

I think it is quicker for somebody to actually say it to you. All that positive written feedback is terrific. One of my goals for this year is to write more. According to (ERO reviewer's name), the whole process of giving children written feedback doesn't take very long and we were talking about marking books and things like that. She said it only takes you a minute but it's doesn't – it doesn't take you just a minute - you might have 8 children in that group, or 8 children you are going to see that day just for that writing session. It is not always that easy. That is the reality.

This is also evident in the statement by Susan regarding formative assessment in Social Studies where she made the point that assessment information gathered in this area has low utility due the New Zealand approach of teaching this subject in blocks of four to six weeks. By the time the information has been gathered the teacher has usually moved on to a new unit in another curriculum area.

6.2.10 Summative Assessment

Frustration has been expressed by teachers over the amount of assessment the school has to carry out to fulfil its requirements to the Ministry, the BOT and ERO. Mary expressed her views on this issue at Endeavour:

There are two levels of assessment. Richard needs it to show the Board what we are doing. The Board needs it to show the Ministry we are doing it to get the funding, those requirements. When he looks at it he is looking for

something different from what I would be looking at. What I would want to give the teacher because all those pretty graphs are all very well, but it's not what we want to know. We want to know about Johnny over here, what level he is spelling at or where are his gaps, and what we can do so that we can support him to help him achieve at a much better level than what he is actually currently doing. That is what it is about - the rest of this is nonsense. I hope it is going to get more manageable.

6.2.11 Summary of Assessment Challenges, Manageability and Utility

Analysis and coding of the interview data has led to the identification of eight issues regarding the implementation of assessment.

1. There is evidence of a challenge regarding the structure of some significant assessment tools such as asTTle Writing which begins in year four and the desire to have whole school assessment.
2. A commitment to whole school assessment can be problematic in that it does not always take account of the developmental stage of the youngest children in the school.
3. The utility of assessment material gathered in the non-core areas of the curriculum is seriously devalued through the practice of teaching these areas in topics of six to eight weeks. This has meant that by the time the assessment has taken place teachers had moved on to a new topic in a different curriculum area.
4. Consistency of data collection requires development in some areas. The need for a consistent approach was raised for Running Records and Numeracy Project assessment as well as matching writing assessments when using both the Exemplars and asTTle.
5. The time required to undertake assessment is problematic. This is especially evident in those assessment activities at Endeavour that require extensive one-to-one processes such as Running Records and Numeracy Project assessment. It is also evident in assessment practices that require extensive data entry and moderation activities as evident in asTTle Writing.
6. The establishment of systems at Endeavour that track individual children in all areas of the curriculum have been complex and time consuming.

7. A challenge exists between what is seen as the ideal of formative assessment and the reality of implementing this approach in a crowded curriculum with classes of over 25 children.
8. A constant challenge exists to find assessment activities that are useful for both improved teaching and meeting the school's requirements to report summative data to the Board, the Ministry and ERO.

6.2.12 Summary of Assessment at Endeavour School

The analysis of interview data and examination of Endeavour's assessment documentation reveal some key elements in terms of how this school established assessment and reporting systems and how they dealt with the associated challenges this presented. This analysis also examines how they have implemented the most recent developments in assessment and reports on the implications of this for the manageability and utility of assessment.

Although not formally documented, it is evident that the changes the school has made in assessment over the last ten years have been directed toward meeting eight goals that appear to be shared by the four staff members interviewed.

Assessment procedures have been introduced in two distinct methods. Firstly, a deliberate and planned programme to implement a new assessment practices. This is shown by the introduction of asTTle Writing in 2004 which has dominated the school assessment activity in the last two years and introduced the school to the concept of formative assessment. The second method which involved less direction and certainty is illustrated in the less successful introduction of the Numeracy Project Assessment and the current transition in the assessment of the older children's reading.

While the nature of the research question has to some extent directed the analysis to highlight the more challenging and problematic aspects of Endeavour's assessment activities, it is important to conclude that the four staff interviewed were largely positive about their assessment procedures. Where there were challenges, they were optimistic that these could be met and there was a common understanding of where the school needed to go.

PostScript Interview Richard May 2009

In May 2008 a postscript interview was conducted with Richard. *The Assessment Goals and Challenges in Assessment* which attempted to summarise assessment practices at Endeavour School in 2007 were presented to him to comment on in terms of whether he believed that they were an accurate summary of assessment at Endeavour at the time the research was carried out. In this interview Richard confirmed the goals and themes were an accurate description of assessment practices and issues at the school during that period. While the purpose of the postscript interview was not to gather additional data Richard reported that major changes had been made to the School's assessment procedures since 2007. These changes included the full implementation of asTTle Reading twice a year. STAR Reading had also been retained and administered twice a year. In his 2006 interview, Richard reported some concerns with mathematics assessment which in his words "were a bit light". In 2008 the school had worked with a mathematics adviser and revised its use of the Numeracy Project assessment. It was now using this information for reporting to the BOT. Richard reported that he now had more confidence and understanding in the use of this information. As a result of these developments the school had dropped use of their basic facts tests.

6.3 Case Study Two: Tasman School

6.3.1 Description of Tasman School

This case study is reported using a similar but not identical structure as provided for Endeavour School. As stated at the beginning of the chapter apart from one teacher, all staff at Tasman School had joined the school within the last three years. The only teacher who had been working at the school for more than three years declined an invitation to be interviewed for this research. It was therefore decided to interview only the three management staff. In order to get a comprehensive picture of assessment at Tasman, a more detailed examination of ERO Reports was undertaken than at the other two case study schools.

Tasman school is an inner city school with an ethnically diverse roll. It is a decile 7 school and a number of its parents work in tertiary education. The high staff turnover has a strong influence over all operations at the school particularly assessment.

The school had a roll generated staff of seven teachers but there were 13 teachers employed at the school. The staffing included an attached special unit. The school was organised into two syndicates of four classes. The Junior Syndicate was made up of a new entrants' class, a year one-two, two-three and three-four class. The Senior Syndicate had a similar structure of four composite classes: four-five, five-six, six-seven, seven-eight. The Deputy Principal led the Junior Syndicate and the Assistant Principal led the Senior Syndicate.

The senior management team of the principal, assistant principal and deputy principal led most decision making regarding assessment in consultation with the rest of the teaching staff.

Table 6.3: Tasman School Description (2005 ERO Report)

Roll	
Decile	7
Structure	Full primary
Ethnicity	NZ European/Pakeha 52%, Asian 15% Maori 12% African 12% Samoan 6% Other 6%
Principal	Brian
Deputy principal	Judith
Assistant principal	Ross

Table 6.4: Assessment and Reporting at Tasman School February 2007

ATOL Contract	Yes – but most staff have changed since then
Literacy Contract	2005/2006 reading especially ESOL
PAT Tests	No (discontinued 2005/2006)
STAR	Yes, twice a year
asTTle Writing	Start 2007
asTTle Reading	Term one and four
asTTle Maths	Term one and four
Numeracy Project	Full diagnostic at school entry after Year 2 Gloss – twice a year
The Exemplars	Writing in Junior School, informal class use
Spelling	Peters
Basic Facts	School developed
Running Record	Twice a term up to year 3, then it is unclear what is expected
ARB	Informal use
SEA	Not maths, use diagnostic instead, some taken informally
Six Year Net	Yes
NEMP Tasks	No
Cumulative Cards	No
Portfolios	Juniors – Profiles sent home Middle and seniors – termed assessment books go home every term
Tracking and monitoring books	Social Studies and Science based on AOs, <i>in transition to key competencies</i>
Student attitude surveys	No
Written Report	Reports at the end of the year, Assessment books go home three times a year
Parent Interviews	Three times a year start with goal setting
School Management System	MUSAC

6.3.2 Assessment Practices, February 2007

Current assessment systems at Tasman School have responded to high staff turnover, including a total change in management staff during the period June 2005 to July 2006, and a challenging ERO Review. These two factors and the introduction of new assessment procedures give a particular focus to assessment at Tasman School. The school introduced asTTle Reading as a result of a Professional Development contract and at the time of the study was planning to introduce asTTle Writing. STAR Reading

has also been introduced into the school recently along with asTTle Mathematics. At the same time the school recently discontinued its previous extensive use of PAT Tests. The school also revised its methods of tracking individual achievement in response to ERO comments. In addition to these changes, Tasman School was in the process of introducing increased electronic tracking of achievement through its new school management software. This project was being led by the assistant principal. There was an increased emphasis on formative assessment and ongoing discussion on the purpose of individual assessment procedures. While this context sounds daunting, the overall impression gained from the management staff was that they were optimistic about the changes they had introduced.

6.3.3 Seven Major Themes in Assessment at Tasman School

The rapid change in practice at Tasman School highlighted seven major themes in assessment:

1. Seeking stability after a period of rapid change
2. Aiming for consistency of practice
3. Clarifying the purpose of assessment
4. Creating systems for tracking individual achievement
5. Electronic recording as a driver of changing assessment practice
6. Meeting the needs of the community
7. The importance of the recent professional development contract and the use of asTTle Reading in developing practice in formative assessment.

Seeking Stability after a Period of Rapid Change

The year 2006 saw a great deal of change in Tasman's assessment procedures. As Judith stated: "They [the assessment changes] have been absolutely quite huge really. Especially this year and I think this year has been the year of transition."

While changes in staff did present problems it did have positive advantages in introducing some large changes in direction. In responding to a question about the advantage of having the deputy principal and principal both new to the school, Judith stated:

It did help and also we came in and we discussed where we would like to see it go so we have that objective to aim for. The other thing is there is only one staff member left. So there is nobody around to bring up what was done before which makes it easier for us able to change the system.

Having been through a period of rapid change the school was seeking to stabilize its procedures. This search for stability was most strongly articulated by the principal Brian. The three quotes below give his perspective.

Well last year we put everything in a draft fashion for a review. This year we have just reviewed it now for term One and term Four testing. I think for next year we have actually got it right. I think this year we have over-assessed. Last year we didn't know if we were PATing or asTTling but I think for next year we have got it.

No I think there won't be many changes in the next 12 months. I think I would like to see it run now.

I guess what's working well is that we have got a routine twice a year. That's taken 12 months for everyone to get up to speed.

Ross commented on the changes, which although he saw as an improvement had still taken up a lot of time:

Probably how often it is, how much there is. Sometimes you feel you are just testing and you don't actually teach and move them. This year we have gone to doing a lot in term one and it was frustrating. We did a lot of standardised testing and we are going to do that again at the beginning of term four. It gives us two terms of getting into them whereas last year we had asTTle term one, STAR term two, asTTle Term Three, STAR Term Four. I just felt like all you were doing was assessing.

Aiming for Consistency of Practice

An indication of the stability that Brian searched for is demonstrated by consistency of practice in assessment procedures. This is an area that Ross, the assistant principal, believed the school had made good progress in:

It is consistent throughout the school. So that's useful everyone does the same stuff. asTTle as a tool gives consistency across everyone. The same with STAR and all the other standardised tests.

Other changes this year have been the big one – moderation consistency between levels – like using the exemplars making sure the person next to me is assessing with the same quality that I am assessing with is slightly concerning – something to be aware of. It is in the PD we have been doing this year.

Clarifying the Purpose of Assessment

The introduction of new people in the assistant and deputy principal positions brought about a culture where the purpose of assessment was debated and challenged. This questioning of the purpose of assessment focused on questions of the manageability and utility of various assessment procedures. As Judith stated:

I wonder if sometimes we assess to assess and that's why I like that idea of asking why we are doing this and developing a culture that does that and I think that's where we are heading and that is what I like. That's why we dropped PATs. Why do a PAT, do a STAR, do an asTTle, do a this, do that. If you are going to add one of those new assessments in you have got to take something out. We are not very good at taking something out are we? Therefore we become overloaded with assessment and I don't think that is the objective.

Ross, who had been originally employed as a scale A teacher, outlined a debate about the use of a spelling test:

Yes like the one I had when I was a scale A and the old DP about Peters Spelling and my basic thing was why, why are we doing this test, and we are still having that discussion and we still do it? There is no analysis of it. So why am I doing it. That seems to be the change in thinking now. If you can't tell me why we are doing it, well then, we are not doing it.

Creating Systems for Tracking Individual Children

The tracking of individual children was an area where the school was challenged by ERO. Brian found this challenge particularly frustrating and has moved rapidly to address the issue:

There were individual profiles but they didn't have a lot in them I think they were introduced in 2001 and I came in last year. I found it very hard to track kids individually. I found the tracking sheets after ERO had gone. So what we had done this year is rewriting for specific learning outcomes as well as spelling results and asTTle. It is basically that all the results we are getting we are recording them electronically. Each child's got a file on MUSAC. It is basically, everything that was in the folder.

As Judith also stated on this issue: "There were other forms of assessment but they were also put into the summary of the levels and there wasn't any individual tracking of children. It was all done in a summary in class sets."

Electronic Recording as a Driver of Change

Related to the search for consistency and greater tracking of individual student achievement, Tasman School has put a lot of energy into electronic recording of student achievement. Ross, who was the main driver of this change, hoped that developments in this area would make assessment more manageable. Judith saw the move to more electronic recording as improving the tracking of student achievement and increased formative assessment:

We are going to go digital and track them that way so that will be quite a big change and teachers will need to be “au fait” with the system and hopefully we are talking about doing more formative assessment, feeding back and focusing on the child and meeting the child’s needs. So that to me is the big change.

Meeting the Needs of the Community

For many years Tasman has been considered a multi-cultural school with a large number of ESOL students. In the 2005 roll, 64% of students were either Pakeha or Maori with Asian, African and Samoan included in the remainder of the roll. Ensuring that assessment procedures met the needs of these learners was a strong focus for Brian:

Yeah I have always thought the PATs were biased and very ambiguous for ESOL children and they were so old fashioned with multi choice questions.

Well I think it has opened people’s eyes up to the fact that a lot of testing is culturally biased and some children are not given enough direction.

I still think that it is just the format of the test – The GLOSS – Testing well that’s one to one so it’s a lot better for ESOL children. I think Running Records are still really pertinent because they are one-to-one. I think it just made people more aware that they have to be more considerate to ESOL children.

Ross was also sensitive to the needs of new learners of English as he described an adaptation to make Numeracy Project testing more manageable: “Last term I did a written test that covered certain things and transferred that to the tick sheets pulling certain kids and doing it orally with the ESOL kids to make sure about the lower groups.”

Recent professional development contract and the use of asTTle Reading

Tasman had recently undertaken a professional development contract based on guided reading. It introduced the use of asTTle Reading into the school and, in Brian's views, made the staff more sensitive in the ways they assess ESOL children. The use of asTTle Reading and its "What Next" function had developed the staff's understanding and use of formative assessment.

6.3.4 Assessment Tools and Procedures

As noted in Table 6.4 Tasman School had introduced a number of new assessment procedures in the previous 18 months.

Numeracy Project

The school had undertaken the Numeracy Project in 2000/2001 which meant only one member of the present staff was at Tasman when the project was undertaken. After debate in 2005-2006, it was decided that the school would carry out a full numeracy assessment of all new entrants and then undertake a GLOSS test twice a year after that. Brian reported that he felt that the ongoing GLOSS and the information the school was gaining from asTTle Maths was sufficient. As principal he was pleased with the information that was gathered from this process. In response to the question: "What about the utility of your Numeracy Project assessment?" He stated:

Absolutely one of our planning targets this year was to get 75% of our children at their age level or above. The students are meeting this and they have got the strategies.

Of the three participants interviewed, Ross who taught year 5 and 6 found the Numeracy Project assessment most challenging. He had missed training in the Numeracy Project through his teaching in England:

We do the GLOSS twice a year and the tick sheet with the different stages early additive etc. The approach is that you tick what they can do. I could never get my head around that as I missed all the training while I was in the UK. So when I came back I did a little short course. This is the first year I have got it up and running and it is the same issue as the Running Records. The problem is the manageability. I know some teachers just do it as they go and if they are working with a group and a kid shows them that they can do something they record it. Ideally that's the way it should be. It is a working document – great you are doing it – you have shown me three days in a row that you can do it. That's fantastic! That's ongoing assessment and I haven't

quite got that going yet because I haven't got that system running yet because I haven't quite figured it out.

asTTle Reading

As stated earlier, the application and use of asTTle Reading had been developed out of a current Literacy Contract. Staff reaction to the Contract and the use of asTTle Reading has been very positive. Judith commented on asTTle Reading:

I actually don't administer asTTle but I have been part of the training and I know it has been quite beneficial in getting information on the children because it breaks information down so efficiently and we know there is a big gap in vocab in the class. Some of that is because there are a lot of ESOL children. With the individual child you can see where they are doing really well. So I like asTTle for that. I know it is beneficial. The only thing that is hard to manage is the data entry.

Brian: "asTTle is working well; everyone has got familiar with how it works, going on line and getting the next steps, that sort of thing. That's taken 12 months for everyone to get up to speed with it."

STAR

In addition to asTTle Reading the school also administered STAR which they described as manageable. Brian, however rated asTTle as a superior test in terms of the information it provided.

Running Records

The school undertook Running Records as another form of assessment in reading. The Price Milburn (PM) Assessment Folder was used to administer a running record for all students until year 3. Beyond that stage it was unclear what the procedures were as all members of the management team had a different understanding of school-wide expectations, including whether seen or unseen texts were to be used. This conclusion is perhaps an understandable reflection of a staff that had experienced a lot of changes in assessment procedures in the 18 months, with many of those changes focusing on standardised tests.

Assessment of Writing

At the time of the interviews (February, 2007) the school was about to begin using asTTle Writing (years 4-8) and The Exemplars (years 1-3) for the assessment of writing.

The Exemplars, ARBs and NEMP Tasks

Judith used The Exemplars in the junior school for writing and had made some use of this tool for art and mathematics. Apart from this there was very little use made of the Exemplars, ARBs and NEMP tasks; any use was of an informal nature and confined to individual classes.

School Entry Assessment

School Entry Assessment is used in the school. However, the original approach has been modified to take account of manageability issues and changes in understanding of the curriculum. The Numeracy Assessment has replaced the mathematics component of the SEA. Judith had attempted to address the manageability of the tool, in particular the fact that it takes up to 45 minutes to carry out one assessment, by trying to incorporate as much of the assessment into the normal class programme. The results are passed on to the principal, but other than that little use was made of them.

Folders of student achievement information

There are two portfolio-type systems used to communicate with parents. Judith was positive about the system in the Junior School:

In the Junior School we do profiles. There is a learning goal. The children then have samples of work and then it is analysed and it goes home to the parents. It works well. They go from these profiles to this book then the assessment book gets put into the portfolios. I think there is a lot more self assessment coming into the books.

Ross describes the system used in the middle and senior levels of the school:

We have a book where they glue work straight in the book and that includes any end of term assessment and end of year stuff. Their final writing assessment they do might do in there, Maths tests get glued in. That goes home each term. It is the home school communication. The Maths tests would have the objectives and you would highlight it if they had achieved it or if they didn't you wouldn't highlight it – you would give them next steps so that their parents know.

Ross believed the system was manageable if the teacher “keeps on top of it.” In terms of utility he saw the system as useful for communicating with parents but not of great value for informing teaching.

Non-core curriculum assessment

The school's approach to assessing in the non-core curriculum areas appears to be in transition. Previous ERO reports had challenged the school for a lack of assessment against the levels of the National Curriculum. There was an impression, although not directly stated, that the current management team had inherited a system without enthusiasm and were looking forward to using the key competencies of the revised curriculum to implement change. The impression was also given that attention to other aspects of assessment has prevented the staff from focusing on changes in this area.

Ross outlined some issues around assessment in these areas:

Traditionally we have done our units and assessed them using mainly knowledge based assessment. Such as do they understand the water cycle? Ticking yes, no or working towards. I always had doubts about what children need to know. I don't think the knowledge of these subjects is overly crucial.

For kids with skills it doesn't really matter, if a kid knows the rain cycle. I would rather know that they had the ability to locate information and find it. They could find anything. I would rather be assessing skills than knowledge.

Formative Assessment

All three staff expressed enthusiasm for the use of formative assessment. As Ross stated:

Formative assessment, I guess it helps me with what I need to teach, where I need to go, what the gaps are. So it's done at the start. Which is what I like about asTTle. They are giving you things to work on instead of at the end finding out that they do not know. The kids quite like it when we start with me saying this test is for me, don't worry so you can get some wrong, the more wrong the better, so we have got more to work on. So I like that. It shows up a lot of the gaps, what things we need to work on. Every writing sample we see we go over it and talk about the what next and whenever they write we conference them, we write it in their what next log.

However, Brian outlined problems that existed between the ideal as articulated by ERO and the reality in its implementation.

It's really only when it gets down to the nuts and bolts of formative assessment. Verbal feedback – which is better value for the ESOL children. We have been working on mainly verbal feedback. ERO challenged us on this. They thought there is not enough written feedback. Here, especially with ESOL children, talking to them and smiling and pointing out what they have to do works. I suppose there has to be a balance ...

I mentioned feedback before; I guess written feedback has its levels of quality. That's something that needs to be very consistent but that it is very time consuming and I've tried it. I sat down in a handwriting lesson in a junior class and ERO had been mouthing off to us that you have got to write down what the next steps are for the children. I sat down and what use to take me as a teacher, 20 minutes a lesson I sat down and wrote them written feedback. I had a queue out the door.

6.3.5 The Manageability and Utility of Assessment

The three participants made a number of comments on the manageability and utility of assessment. In respect of the Numeracy Project, Ross commented on the challenge of carrying out these one-to-one assessments "it is the same issue as the Running Records, the problem is the manageability"

And the same person on assessment in general:

No parts of it are manageable, parts of it are useful. The big picture is there is too much to assess too often. What is the purpose of some of it? All the assessment we do, I don't think it has enough impact on the kids. I have got 100% effort I can put in. If I am putting 50% into assessing I am only putting 50% into planning and teaching. Certain times, the end of term is very stressful. There could be less and not impact negatively on children. I know there are certain things that have to go to the Ministry, but as a classroom teacher a lot of assessment I do doesn't impact on children.

On a similar theme Judith commented:

It is frustrating that there is so much assessment. There is lots of it. That you are required to do so much of it. Sometimes the teaching doesn't actually happen. What I would like to see is assessment so integrated, so embedded, that people wouldn't even know it was assessment. Rather than doing a STAR here and then you analyse it and I think all those things are good – I am not saying they are bad but in an ideal world it should be so integrated it is not different from what is happening.

The message is very clear from these people: while some of the assessment is useful, manageability overall is a problem. There needs to be much less assessment and better integration into teaching.

6.3.6 ERO Reviews

An examination of ERO Reviews at Tasman School illustrated a number of key issues in assessment procedures at the school as well as the Review Office's own developing understanding of assessment. Since the advent of ERO, Tasman has had six reviews:

1993 (report not available), 1994, 1996, 1999, 2002 and 2005. When the five reviews since 1994 are considered in total, the Review Office frequently reported the school needed to make considerable development in its assessment procedures. The reviews also highlighted the frequent changes the school has had in its management staff.

Effectiveness Review September 1994

The assessment elements of the 1994 Report highlighted three major themes:

1. The need to be able to demonstrate that the school was making a difference to student achievement over time: “Like many schools at this time the Board is unable to demonstrate it is making a difference to student achievement.”
2. The lack of school wide systems leading to the recommendation: “the school considers reviewing and evaluating current assessment policies and procedures to provide school wide information about student achievement.”
3. The view that learning objectives are central to effective teaching: “the lack of learning objectives in most curriculum areas makes assessment and evaluation difficult.”

In addition to these three points the Report made a number of other comments on assessment:

A range of useful strategies has been initiated to record individual achievement, including student files and some checklists, particularly in the junior school.

The quality of entries in individual files in the middle and senior school is generally poor. Assessments are infrequently recorded and a number of curriculum areas do not have any achievement information recorded.

While several samples of written work are often included in student files, these are rarely analysed and supported by professional comment. In some cases samples of work are not dated.

ERO noted the use of a number of assessment tools:

- Running Records
- PAT
- Maths individual records kept
- Formal Curriculum mark book
- Science assessment against objectives.

December 1995

The December 1995 Report noted that the school had worked to address assessment issues raised in their previous report:

Improving monitoring and assessment systems was one of the major recommendations made in the previous ERO Report. The school invited Teacher Support Services advisers to survey the school's needs and prepare workshops for the staff. This resulted in an extensive trial of standards based assessment and the introduction of useful files of individual student's work. Progress cards and roll books contain specific information on student achievement.

In the 1995 ERO report, a major theme was the use of curriculum objectives for teaching and assessment:

There is a continuing need to include specific, measurable objectives derived from broad curriculum objectives in all planning. These objectives could also be used for evaluating programmes.

Ensure teachers use specific measurable objectives for planning and evaluating programmes.

The Report also commented on reporting on achievement to the Board:

Staff will be making further decisions in 1996 about procedures for planning, assessment and recording. This should enable the school to collate relevant achievement data, and inform the Board about trends and issues in the curriculum.

March 1999

The 1999 Review began with again acknowledging the progress the school had made and noted the changes in senior staff.

Considerable progress has been made to address the actions and recommendations in the 1999 ERO report. This includes strengthening curriculum management to guide programme planning and reviewing the assessment systems to meet national guidelines. Since that time there have been many changes in personnel, including all senior management positions.

The 1999 Report for Tasman returned to a number of themes noted in the earlier reports using slightly different language:

No information on monitoring against the national curriculum levels (also stated in the 1994 and 1995 Reports).

Developing a school-wide assessment programme that meets the requirements of NAG 1 and that strengthens the identification of barriers to learning (also stated in the 1994 Report).

August 2002

Since the 1999 Review, the school had appointed new people to all three senior management positions. This Report was arguably the most positive the school had received. In the area of assessment and evaluation the Report stated:

The specific focus of teacher planning required for the early numeracy project provides a valid basis for monitoring and assessing student achievement. Achievement information gathered during this development is of good quality.

Reading and writing progress is monitored at regular intervals using standardised tests and the school's writing exemplars to assist with teacher moderation. Some teachers, as a result of sound formative assessment, record constructive written comments about student progress and identify next teaching and learning steps.

The 2002 ERO Report mentioned the use of standardised tests in literacy and numeracy that had not been covered in detail in other reports. The 2002 Report is also notable for the first use of the terms formative assessment and teacher moderation in the Tasman School review. The review team made a number of comments on target setting and expectations:

Expectations for student achievement are broad and structured to be monitored twice over eight years. Targets for monitoring student progress and achievement at more regular intervals do not support these expectations. The development of specific targets would provide the board with a measure against which student progress can be reported.

The theme of the use of national achievement objectives that was mentioned in previous Reports was noted again:

Some teaching practice does not clearly scaffold student learning in reading and writing. Where this is apparent, programme planning does not always break down the national achievement objectives into clear, specific and measurable learning outcomes for the development of a teaching sequence and valid assessment.

It is also notable that the Report returned to a theme from the 1994 Review that had concerns about the lack of student achievement over time.

October 2005

Three years after the 2002 Review, Tasman had experienced a great deal of staff turnover with only one staff member still working at the school from the teachers who were employed there at the time of the last ERO visit. The 2005 Report acknowledged further professional development in both formative assessment strategies in the teaching of writing and developments in assessing against the levels of the curriculum in the strand of English. The issue of tracking student achievement over time raised in the 1994 and 2002 Reports was raised again: “In addition, the school has identified a need to track student achievement over time, as part of its ongoing review of the quality of teaching.”

While the 2002 Report introduced the use of the term formative assessment a strong theme of the 2005 Report involved the collection of evidence and analysis of achievement data:

Following the 2002 review, the staff undertook professional development in formative assessment strategies through the writing programme. These are now well established in teaching practice across the school. The school has begun to collect achievement data in writing against curriculum levels and this review recommends that board, senior managers and teachers further analyse data in reading and writing to plan, both at the class and board levels, for the diverse needs of students.

The board and staff have yet to develop a shared understanding of evidence-based, self-review processes that link the analysis of student achievement, the quality of teaching and the school’s strategic goals at both the classroom and board level.

With regard to formative assessment it was obvious that the Review Office held the view that feedback to students should be carried out in writing: “While students receive good oral feedback about their learning, written feedback with next steps for individual or groups of students are not recorded in workbooks.”

The 2005 Report investigated the Health and Physical Education Curriculum as an area of government interest. The Review’s comments regarding assessment in this area in many ways illustrated ERO’s views on assessment in the non-core areas:

Further issues to consider for a comprehensive and useful review of physical education and health include the use of assessment strategies other than teacher observation, such as the exemplars; or National Education Monitoring Project tasks, to determine how well students are participating and learning.

An implication from this comment was that schools in general should be assessing formally all seven areas of the New Zealand Curriculum on an annual basis and using external tools.

The final recommendations of the 2005 ERO Report were important for this case study as they set the context for the staff interviews that were carried out in 2006 and 2007. Recommendations included:

The board of trustees and ERO agreed that:

- 1.1 *The school community, led by the board and senior managers, will undertake an ongoing, evidence-based self review of the strategic direction, to establish and maintain common understandings of priorities for the children of the school. This will be reflected in clear documentation and practice.*
- 1.2 *Senior managers and staff will continue to develop shared pedagogical understanding of effective teaching and learning in literacy to meet the diverse needs of students through:*
 - *in-depth analysis of student achievement;*
 - *links to the curriculum;*
 - *use of effective teaching practices;*
 - *scaffolding learning; and*
 - *ongoing formative assessment.*

The school's reaction to the 2005 ERO Report

It was interesting to note the reaction of senior staff of Tasman School, all of whom were new to the school, to the 2005 Report. While the Report was challenging, if read in the context of the previous ERO Reports the statements on assessment were, with the exception of 2002 Report, one of the best the school had received. Tasman School had gone through a large number of management staff in eleven years, including four principals, and in the cumulative views of a number of review teams, always struggled in assessment.

It would appear from the comments of the senior management that they took ERO's comments seriously and were motivated to address the issues raised. As noted by Brian:

When I came in last year basically they [ERO] took me to the cleaners. I had only just found the filing system. I had only been there six months and had not really decided on the direction of the school and I guess that's what they were looking for. They came in and expected, I think, a lot more and some written feedback; in some places they were correct.

Ross, who was a scale A teacher at the time of the ERO Report but had subsequently been made AP noted:

The comment from ERO was how I would know how this kid is doing. And we have had to put something into place. Their other point was we used to write each learning step and the end of a unit as part of writing, next we will work on paragraphs. We got rid of that because they said it is pointless. So now it's sort of ongoing. We just got rid of it at the end of the unit.

Judith noted:

ERO arrived the same week I arrived. I think they highlighted some issues about tracking children and that has been some of the reasons for the change. So yes, in that sense we have changed because of some thing they have said and you could say that they have influenced what we do in our practice.

6.3.7 Summary of Assessment at Tasman School

In summary, the study of assessment at Tasman School illustrated a highly motivated management team that was in the process of implementing substantial change to the school's assessment procedures. Such rapid change did present some problems regarding a shared understanding of procedures and the successful embedding of new approaches in a sustainable manner. In terms of the manageability of assessment it was noted that a major driver of assessment change in Tasman School was addressing a challenging ERO review. There may be a concern that this issue could lead the school to adopting assessment practices that were challenging in terms of manageability. It was noted earlier in the chapter that Tasman was undertaking four formal reading tests a year that may provide manageability challenges.

Postscript Interview May 2009

Brian was interviewed again in May 2009. The goal of this postscript interview was to seek verification that the researcher's summary of assessment practices and issues at Tasman School in 2007 was an accurate description. Brian confirmed the *seven major themes of assessment at Tasman School* as an accurate description of the school in 2007. Although the purpose of this interview was not to gather new data Brian reports of subsequent changes in assessment practice are worth noting, Brian reported that the

major change in assessment practices since 2007 was the introduction of asTTle Writing, which was going very well. Brian reported more confidence about assessment practices as a result of a successful ERO review in 2008. He commented that the fact that the leadership team had now been together for five years had brought a sense of stability and confidence to what they did at Tasman in all areas, particularly assessment. Brian reported that a new BOT had improved relationships between the BOT and the staff. However, he noted that as a result of the composition of the new BOT he was reporting a lot more formal achievement data to the BOT. He stated “they are the sort of people who love that stuff”. However, he did not see this as problematic as the improved relationship and trust outweighed manageability concerns in this area.

6.4 Case Study Three: Kupe School

6.4.1 Description of Kupe School

Kupe School is a decile nine school (the highest of the three case study schools). It has a roll of close to 300 students, the majority of whom are classified as New Zealand European. Both the principal and deputy principal have been at the school for over ten years. The school has low staff turnover and has experienced a series of positive ERO reports. Kupe is the only contributing school of the three in the research.

The school had a teaching staff of 17 which was made up of a principal and deputy principal without classroom responsibilities. There were 11 classrooms in operation with an additional new entrant class being established in the second half of the year as roll numbers grew.

The school was organised into three syndicates. The year five and six syndicate consisted of there year five-six classes. There was a similar structure at year three-four with three composite classes. Both these syndicates were led by a senior teacher. The Junior Syndicate was led by the assistant principal who taught a new entrant class. There were four other classes in this syndicate at the start of the year. These were year one-two composite classes.

The senior management included the principal, deputy principal the two senior teachers responsible for the leadership of syndicates and the assistant principal.

This case study involved semi-structured interviews with four staff members at Kupe School: Rebecca the principal, who had been at Kupe for 10 years, Maria a 6th year teacher who was leader of the year one-two syndicate, Jill a very experienced scale A teacher who has been at Kupe for 10 years, and finally Ruth a year five-six teacher who was in her 3rd year of teaching. Rebecca, the principal, was interviewed in November 2006 and the other three teachers in May 2007.

Table 6.5: Kupe School Description

Roll	278
Decile	9
Structure	Contributing
Ethnicity	NZ European/Pakeha 73%, Maori 12% Other 14%
Principal	Rebecca
Syndicate leader	Maria
Scale A	Ruth
Scale A	Jill

Table 6.6: Assessment Practices as at Kupe School as at December 2006

Tool or Practice	Description
ATOL	2004/2005
PATS	Listening continuing
<ul style="list-style-type: none"> • Listening Comprehension • Maths 	Reading Comp and Vocab discontinued after 2005, Maths after 2006
STAR	April /November, since 2005
asTTle Writing	Since 2004 – came in as part of AtoL
AsTTle Reading	No
asTTle Maths	No
Numeracy Project	Began Numeracy Project 2006

Basic Facts	Yes
Spelling	Daniel and Diak
The Exemplars	Writing Junior School; carried out throughout the school before AsTTLe writing, some informal uses in other classes
ARBs	Some teachers, very informally
Running Records	Used in Junior school – policy in transition in senior school
New Entrant Assessment	Partial
NEMP	No
Portfolios	Very comprehensive system
Social Studies, Science	Achievement Objectives – analysed, gender, ethnicity
Essential Skills	No
School management system	E Tap

6.4.2 Assessment Themes

An examination of the four interviews and supplementary document examination at Kupe School identified six major themes regarding assessment at the school:

1. Assessment activities should inform teaching and learning.
2. Understanding of the Ministry of Education's and Education Review Office's thinking on assessment.
3. Collaborative work with other schools on assessment.
4. Assessment in a period of transition.
5. The dominance of the use of portfolios in teachers' discussion of assessment.
6. A very comprehensive assessment system.

Assessment activities should inform teaching and learning

Kupe has introduced a very comprehensive and ambitious assessment system. To fully understand the school's assessment system it is necessary to begin by exploring in some detail Rebecca's philosophy and approach to teaching, planning and assessment. Rebecca began at Kupe in 1997 after working as a DP in another school

as well as a period of two years in the early nineties as a school adviser, assisting schools in the implementation of the English curriculum.

Rebecca reported her shock on returning to teaching her year 4 class after two years as an adviser. While she had been advising teachers about assessment strategies for the English curriculum, the full impact hit her when she had to confront all the assessment strategies for six other curriculum areas:

It was the biggest grounding I had ever had because English was one thing and there was all this other stuff It was all those assessments. The new curriculum was introduced at the same time as ERO. So we were checked for compliance in the midst of a whole lot of new learning.

By 1997 five of the seven new curriculum statements had been introduced in schools: Mathematics (1992); Science (1993); English (1993); Technology (1995) and Social Studies (1997). This rapid change in curriculum plus the increased assessment demands had placed a lot of pressure on schools. Rebecca described her experience of assessment on arriving at Kupe in 1997:

My experience was that a lot of it was done for compliance and not being derogatory but there had been a huge amount of professional development but nothing had changed in terms of what was happening in classrooms and people were gathering information and there was a real anti about planning: we are planning for ERO, or we are planning for the Ministry, or planning for the principal; and we have to have this assessment for ERO or assessing for the principal, or assessment for the Ministry.

What was happening was reasonable but people didn't have ownership of it, nor understanding, nor did it translate into what was happening in classrooms. The teachers were the gatherers. It was mainly PATs, basic facts, Running Records which were variable in terms of how often and when but it was this week and this term you do this and this week and this term you do that.

From this position, Rebecca and her Deputy Principal, who was appointed shortly after Rebecca's arrival at Kupe, set about clarifying their positions on the relationships between planning, teaching and assessment. As noted by Rebecca:

We needed to get some common understanding of what we understand as effective teaching and learning and we brought in (names professional consultant). We went right back to the basics what is effective learning and teaching what does it mean for you and you and you? We came out of that with some principles really that underpinned what we did. So we said we have got to know where kids are at. So how do we find that out? What tools do we use.

The philosophy developed in this period provided the base for the school's approach to planning, teaching and assessment activities ten years on. Kupe aimed to use the same assessment tools for informing teaching, reporting to parents and the BOT and the MoE planning and reporting requirements.

I think the key feature is the only assessment we do is what is going to inform teaching and learning. We use the same information to inform the BOT about achievement and the same information for planning and reporting. We don't use a lot of things for different things but use the same data for a lot of different purposes and report it differently.

I think our assessment is working well in terms of informing our planning. We don't use two year plans for anything now. We don't use two year plans for what we teach. Any decision on planning is based on the information we have about the kids and that will be from the formative stuff; the day-to-day stuff as well as the long term, norm referenced stuff.

The initial changes the school made were not without challenges from staff: It was the understanding how assessment could inform learning and teaching. I wasn't strongly compliance driven and some of the teaching team at that time were quite bitter. There had been some disharmony in the staff with the BOT prior to me coming and there was a bit of an overhang with that. There was a bit of "this is another thing we have to do" and teachers were tired with the new curriculum stuff. It was another thing "you mean to say we have to plan and the planning has to be informed by what the kids need – how are we are going to that, we need to have a life."

Rebecca's description gave an historical perspective on assessment and teaching practice in the decade of the nineties. From her view the rapid changes associated with the introduction of the new curriculum and associated assessment had led to teachers undergoing a high level of activity for, in her view, little purpose. Rebecca's response was to develop a shared understanding of what were the key elements of good teaching and learning and base assessment activities on that understanding. It is interesting to examine how this philosophy has translated into assessment practice ten years later and what this meant in terms of manageability and utility.

The desire to use one form of assessment for all purposes of assessment was supported by Maria the year 1-2 syndicate leader:

So I don't see any point in doing three different things to assess one thing. I am looking for one form of assessing to give me information for my own teaching, to give Rebecca the information, to give the BOT – whatever, whoever needs that information. I think carefully about what I am going to do to get that multi purpose information out of it instead of putting your kids through heaps amount of drama.

Understanding of the MoE and ERO thinking on assessment

Rebecca's previous experience as adviser in the implementation of the English curriculum, her high level of professional reading, and contacts with other leading professionals appears to have given her a thorough understanding of both the Ministry of Education and ERO perspectives on assessment. In her interview she listed the influence of her Masters degree, earlier Advanced Studies for Teachers (ASTU) papers on assessment, the writings of John Hattie and Brian Annan, as well as visiting the MoE's *Leadspace* website at least twice a week as influences on her understanding of assessment. In addition to this, Rebecca and two other local principals initiated their own joint initiative across the three schools looking at writing achievement and developing their own exemplars.

Rebecca felt ERO has not influenced the school's assessment practice. Kupe School has received a number of very positive ERO reports which could allow a conclusion that they are a reflection of the principal's very informed understanding of current and past intentions of government agencies for assessment practice.

An analysis of the ERO Reports of Kupe School from 1991 shows a continuous record of positive reviews. The 1996 Report did express a need for management to develop further collation and analysis of student achievement data. However, this review did go on to state "A guideline paper on monitoring student progress and reporting to parents is an exemplary model for meeting both these requirements."

The three Reviews under Rebecca's management have been extremely positive in the area of assessment strategies. The 1999 Report stated:

The school-wide assessment programme provides valid and reliable information. A full range of assessment procedures and criteria is developed for multi-level units. Aggregated data is now available to use as information for better learning. Portfolio collections of student work are important features of the school-wide programme and the portfolio conference process includes sharing work regularly with parents and including parents in the goal setting process.

Both the 2003 and 2006 reviews were equally positive in their evaluation of Kupe's assessment procedures.

Collaborative work with other schools on assessment

As mentioned above, in 2000 Kupe decided to work with three other schools in developing their understanding of assessment in writing. Rebecca described the process:

We started in maybe 2000, our school worked with [school name] and [school name] that came out of the Literacy Leadership programme where we said we think our kids are doing OK in writing but how do we know? So we worked together and we developed exemplars for writing and this was before the Ministry exemplars came out.

In the last two years the school had been involved with a local MoE initiative that involved all schools in the local cluster assessing reading and more recently writing together. Teachers have found this a stimulating activity. As Ruth noted: "... we did a session on moderating in our cluster's writing so there were 50 pieces of writing between us, and that really got me knowing and understanding what makes the different levels."

Assessment in a period of transition

The school has carried out PAT testing in reading comprehension and vocabulary for a number of years. This was in its last year as the school planned to replace these tests with asTTle Reading. The school was in its second year of the Numeracy Project training which has introduced a new assessment regime in mathematics.

The school had recently moved from selecting a sample of the child's work for each curriculum area once a term for placement in the student portfolio, to a system where rather than covering all teaching subjects, teachers selected what was most appropriate. This change featured strongly in the interviews with teachers.

The dominance of the use of portfolios in teachers' discussion of assessment

The use of student portfolios occupied the largest amount of time in the teacher interviews. This was understandable as it was the largest element of assessment activity in the school. As stated above there had been a recent change in expectations about the student portfolio. Until recently teachers were required to include a sample of children's work from all seven curriculum areas in the portfolio each term. This

had recently changed so that teachers, and to a certain extent students, chose which work went into the portfolio.

The use of the portfolios is described by Rebecca:

Yes, they go home twice a year with the reports and with a summary report at the end of the year. So they go home with goal setting and they go home with the written report. Parents can come in any time and look at them. It is mandated that they are used once a week as part of the classroom programme. So they are not sitting on a shelf over there. And I mandated that because that wasn't happening. And teachers were saying we have got to do this for the portfolios and I realized that was the compliance thing and if you are doing this for the portfolios it can go in the bloody river because that's not what it is about.

Teachers wrote an annotation for each sample entry. Maria estimated that she would be making about eight entries per term and Ruth six. When questioned, the class teachers reported that they saw the system as being mainly concerned with informing parents. The teachers were questioned about the use they made of the assessment to inform their teaching. Both Jill and Ruth reported that they did not as a rule read the portfolios of their new class at the start of the year.

When asked if she found the portfolios helpful to her teaching Rebecca responded: "For me it is more like a snap shot of what they were doing at the time so it is just basically something from their work that is taken out and put in their portfolios. So no it wouldn't, no."

Jill, who had been at Kupe 10 years, identified portfolios as an example of assessment increasing:

So if I think about compared to when I was first here there is far more assessment now than then. I think some of that is related to having portfolios going home. We stopped doing a report in the middle of the year. When we started doing portfolios at the end of each unit, whatever the unit was, you would do some form of evaluation and you put something into the portfolio. That may not have been what was suggested but that was what everyone took away from the discussion about it.

The previous regime of portfolios had also been questioned in terms of relevancy. As Ruth noted:

I do now, it has been put back on our shoulders because before that it was how we are going to find something for every subject and some I don't see as important. I mean knowing how many times a child can bounce a ball with one hand and with the other I don't think parents would see that as being important.

However, Ruth does express a positive view about the portfolios:

The portfolios even though for me it is about reporting to parents I still see that as valuable. The parents value them. A lot of the conversations with parents on camp say that the kids get very excited about them and come home and discuss them. We have got goal setting coming up shortly so we have to go through and look at how well they might have achieved their goals or what they might need to keep working on or whatever.

A very comprehensive assessment system

Kupe has a very comprehensive system of assessment. It includes the extensive use of portfolios, the retention of the use of individual student (green) cards, and assessment books that report achievement in all areas of the curriculum using the format for levels of the curriculum as used in asTTle. The school also has a comprehensive system of formal testing that in 2006 included PAT Reading (Comprehension and Vocabulary), Listening, and Mathematics as well STAR, asTTle Writing and now the Numeracy Project.

6.4.3 Individual Tools and Procedures at Kupe School

Running Records

The use of Running Records in the school appears to be in transition. There was differing understanding among staff on how often the school expected them to be carried out. Rebecca's view was that after students were reading beyond Gold level, they were given at least three Running Records a year. There had also been recent discussion on the use of seen versus unseen text that did not seem to have been resolved. As Maria, who teaches year 2, put it:

We used to use the PM Benchmarks on unseen texts. Now what I do is for the struggling readers, or the readers at risk, or readers who are still working on the learning to read programme type aspects of things, I use known text. For the ones who are doing quite well, that is the ones who are developing comprehension skills. I will use unseen for them.

Running records are undertaken either during class time or release time. Ruth, a third year teacher of year 5-6 students stated: “the running record is hard work because of the times it takes to do them and the analysing of it takes time.”

The Assessment of Writing

While the use of asTTle Writing (years 4-6) and The Exemplars (years 1-3) came out of an ATOL Contract in 2004-2005, the school had done some exploratory work with other schools earlier. All three of the classroom teachers spoke positively about writing assessment. Jill’s quote is typical of feedback in this area:

One of the things that works well for me is the assessment of writing. As we have gone on and done more work related to that, we are getting a better understanding of the criteria. This means understanding what you are looking at and assessing, and then sitting down with a colleague and sharing each other’s work and writing up where the children are.

There was frustration recorded that, because asTTle does not operate below year 4, the school needed to use two tools.

Numeracy Project

The Numeracy Project began at Kupe in 2006 and the teachers were still involved in professional development. The programme began with all teachers being given two release days to undertake the full diagnostic survey with their children. The interviews revealed little comment on the utility of these assessments, either positive or negative, and judgments in this area may need to wait until the programme has been running longer at the school. However, there were some issues raised about the manageability of the assessment. Ruth who teaches year 5 and 6 stated:

The only problem with assessment is the time it takes. The Numeracy ones are the hardest but you get a bit of release time to do that. It takes 15 minutes for each child. It could be longer. It is pressured. I don’t find it manageable. When I say that, “yes” it is manageable, but there is manageable and manageable.

As a principal, Rebecca found the assessment a challenge to understand through her lack of familiarity with the approach:

I go to all the sessions but it is the first time I feel out of touch with the curriculum because I am not doing it. I guess I could have gone into a class or taken a group; done something about it. I haven’t I find when I am reading the reports now I don’t have a good understanding of what it is saying.

Maria, who was the curriculum leader for the Numeracy Project, expressed some concerns over the reliability of assessment in this area. After teachers had completed

assessments, the results were entered into a national database. The interview with Maria exposed the fact that a number of staff at Kupe were lacking either the skill or motivation to enter assessment data electronically. This could present a challenge for Kupe in a number of areas in the future.

STAR

The STAR Test was introduced into the school in 2005 and was administered twice a year. The test was used collectively with other local schools to gather cluster information. The reactions of teachers to the test are largely positive. As Maria noted:

I find STAR quite helpful. I use it as a supplementary. It gives you the information on things like vocab and on the rest of it is quite helpful to triangulate from the Running Records and professional judgments.

PAT

In the past Kupe had undertaken a comprehensive PAT assessment programme which was being rolled back. Rebecca stated:

Well, we are not going to do the PAT Reading Comprehension and Vocab because we just thought that was overkill when we are doing STAR. STAR is giving us the information and we will probably move to asTTle Reading to drop down from there.

School Entry Assessment

Kupe used an adapted form of School Entry Assessment with the number element being replaced by Numeracy Project assessment.

The Exemplars, ARBs and NEMP Tasks

Other than writing, only a small amount of informal use is made of the Exemplars. The same is true for the ARBs and there was no record of use of NEMP tasks.

Assessment in the non-core curriculum areas

Interviews with the four staff members helped present a picture of assessment in the non-core curriculum areas. To some extent the assessment activities matched the level of the children being taught and the particular curriculum area. Jill who taught the youngest children reported that with her year 2 children, assessment in physical education and oral language was largely informal with a large element of self

assessment. In other curriculum areas the approach to planning and assessment illustrated, to some extent, the philosophy that Rebecca spoke about at the start of this case study. Units in areas such as social studies and science were planned as much out of the students' existing knowledge and interests as from the achievement objectives of the curriculum statements. This did have the effect of teachers in different classes teaching children of similar or the same age using different assessment approaches and plans. In addition to using the achievement objectives in assessment some use was made of the essential skills and a developing use of the key competencies. However, from the interviews it appeared that this was determined by the choice of the individual teacher rather than a planned school or syndicate-wide intervention.

The approach to teaching in these areas reflected the school's implementation of key features of formative assessment, such as the use of learning intentions, success criteria and elements of self assessment. Individual teachers' approach to formative assessment determined their perspective on the use and timing of assessment during a teaching unit. For some teachers the key point about assessment is the activity which happens during the unit.

The recording of assessment in these areas was complicated and involved multiple recording of the same information. Teachers determined a best fit level using the levels of the curriculum. This information was recorded in the teachers' assessment books with a number of these recordings also being duplicated on the developing computer system. These results were also recorded on the children's individual cumulative cards and their end of year report.

Rebecca had considered discontinuing the use of the cumulative cards:

I had thought about giving those away because I thought they are not really adding to what we are doing. One of the reasons I didn't give it away was because we were EROed this year and they wanted school-wide information and I could get the cards; they were all differentiated. They weren't just levelled and I could get school-wide information across the board.

In addition to the cumulative cards and the assessment books some material from each unit was put in the student profile which will be discussed next.

Planning and Reporting

Since the introduction of the Ministry of Education's planning and reporting requirements, teachers noted that there has been a development in the setting and

ownership of the targets by the whole school. Teachers now feel more involved in what targets are set and their implementation has a more central role in their teaching.

6.4.4 Assessment issues for Kupe

Analysis of the interviews of the four staff at Kupe identified two major issues the school was facing in assessment.

Manageability

As with a number of schools, the manageability of assessment did concern some teachers at Kupe. Maria who had been recently appointed as syndicate leader and the teacher with responsibility for implementing the Numeracy Project saw the issues as working smarter:

I personally have been teaching 7 years which isn't a huge amount of time compared to most teachers but is a significant amount of time to have found smarter ways of doing things and that's what I aim to do

Jill saw a need for more time to process what she has to assess:

...for me it's a matter of time, to give credibility to assessment data and time to assess and evaluate what that actually means. I don't have a clear idea of what might be and what this is, a clear quick idea of what their next steps are going to be.

By the time I have marked it and collated it, and hasn't involved thinking about it along the way, that could be an hour or more,

As principal, Rebecca was concerned with the demands put on teachers:

I think it was OK. We have always tried not to do too much but have a single focus and we built it into staff meeting times. I worried sometimes about teachers coming out to staff meetings and facing meaty stuff after they have done a day's work. That really worries me. We don't have many call back days [professional development days in the school holidays] now because I think teachers need a rest.

Consistency

Both Rebecca and Maria (the numeracy leader) expressed concerns with the need to have consistent approaches to assessment throughout the school. In her work with other schools in the cluster, Rebecca expressed a need for clean data:

Always challenging is the quality of the data. So we are working in the [district name] cluster where we are concentrating on raising achievement.

We need to trust it and the first thing we need to look at is gathering information about achievement. It has highlighted to us the variability in the administration of STAR and the administration of asTTle, which impacts on the data. So we looked at how clean the data was.

6.4.5 Summary of Assessment at Kupe School

Kupe has developed a very comprehensive assessment routine based on a well articulated philosophy that has been endorsed by ERO on three occasions. This investigation did highlight the large amount of time teachers at Kupe spend on the portfolio systems that is seen to be effective in reporting to parents but according to the teachers does not contribute a great deal to informing teaching.

Kupe School's experience with assessment illustrates the changing process in New Zealand education since the implementation of the reforms in 1989. While competition between schools was seen as a key element in improving student achievement, the Ministry is now following an active programme of encouraging schools to co-operate. Fancy (2004a) described this process as moving from competition to interdependence. Kupe has worked with other local schools on the analysis of assessment data both informally through the principal's friendship with other principals and formally through a MoE initiative. Teachers reported that they found this contact with other schools in the area of assessment stimulating.

The influence of asTTle in changing school practice was evident in Kupe School through the change from using a triangular or three tick type of assessment (Hill 2000, 2000a; Knight, 2000; Dixon, 1999) to using the asTTle's terms "basic, proficient and advanced" when assessing against levels of the curriculum. Changes in this area have met some resistance from teachers more familiar with the previous system. This resistance presents an example of an interesting phenomenon. It can be very challenging to change assessment practices that are embedded into school practice. To quote the cliché "old habits die hard"; resistance to change can often override factors of manageability, utility and changes in understanding of professional practice.

Postscript interview Rebecca May 2009

In the same manner as had taken place at Endeavour and Tasman a postscript interview was conducted with the principal of Kupe in May 2009. The research summary of assessment practices at Kupe in 2007 which were described as *seven*

themes of assessment were presented to Rebecca to verify the accuracy of this summary. In the postscript interview Rebecca confirmed the *seven themes of assessment* as an accurate description of assessment issues and practices at the school in 2007. Rebecca commented on subsequent changes in assessment at the School. The major change in assessment procedures had been the full implementation of Numeracy Project assessment. Other changes involved the modification of some assessment procedures to take account of the full implementation of the revised New Zealand Curriculum (MoE, 2007).

CHAPTER SEVEN: SUMMARY AND INTEGRATION

7.1 Introduction

This chapter integrates the findings from the three case studies with the major themes of the literature review in order to provide a coherent understanding of assessment in New Zealand primary schools during the period under investigation. It begins with a brief presentation of the contextual background under which the education reforms took place. This is followed by a discussion of a number of major issues and assessment procedures that have been a challenge for schools during the last 15 years. These issues and influences were covered in detail in Chapter Three. The discussion will seek to explain the phases schools went through in adapting to the initial reforms and subsequent changes in requirements and direction. This is presented graphically in a framework titled *three waves of assessment practice*, which locates schools' assessment procedures within the wider context of government policy and practice. This section will seek to answer research question one:

What were the major influences and challenges to assessment in New Zealand primary schools since 1990?

This discussion will be followed by an examination of five factors that have been identified as influencing change in assessment practice in schools: school leaders, professional development, informal networks with other schools, ERO and the MoE. These five factors will be used in an attempt to answer research question two:

How did three New Zealand primary schools establish assessment and reporting systems during the period 2003-2007 that were manageable and maintained a focus on learning?

The response to this question will be presented through a proposed theoretical framework which seeks to explain how three New Zealand schools have developed assessment systems. This framework, which has been termed *four stages in the implementation of assessment*, seeks to explain internal school processes in the development of assessment.

The chapter then answers the third research question which examines the impact of recent developments in assessment:

To what extent do the staff of three case study schools find the most recent changes in assessment practice/tools manageable and provide useful information?

7.2 What were the Major Influences and Challenges to Assessment in New Zealand Primary Schools since 1990?

Chapter Three of this thesis reviewed government documents, assessment literature and the influence of overseas developments to identify major influences in assessment practice in New Zealand schools and to explore some of the challenges they presented. The major influence on assessment at the start of the reforms was the adoption of an OBE model of education. This model, and particularly the way it was structured in New Zealand, led to four major challenges in the implementation of assessment.

1. The lack of direction given to schools in terms of assessment practices and the initial lack of the provision of assessment tools for the new curriculum.
2. The structure of the curriculum in terms of eight levels and numerous achievement objectives.
3. The lack of distinction in the NAGs between the core curriculum of literacy and numeracy and the other areas of learning.
4. The impact of the Education Review office.

While the OBE model was the major influence on education at the start of the reforms, the adoption of the approach to teaching termed *formative assessment* has been a major influence on New Zealand schools in the last ten years. Other influences have been the authentic or performance assessment model, international assessment and, most recently, evidence-based teaching.

It is helpful to begin this discussion by re-presenting Table 3.2, which gives a chronological overview of assessment practice during the period under discussion.

Table 7.1: Assessment in New Zealand primary schools 1989-2006

Year	Initiative
1989	<ul style="list-style-type: none"> • Tomorrow's Schools' Legislation
1990	<ul style="list-style-type: none"> • ERO carry out their first review
1992	<ul style="list-style-type: none"> • Mathematics in the New Zealand Curriculum
1993	<ul style="list-style-type: none"> • The New Zealand Curriculum Framework • The National Education Guidelines • Science in the New Zealand Curriculum • First Assessment for better learning (ABeL) professional development contracts begin. The name later changes to Assessment to Learn (AtoL)
1994	<ul style="list-style-type: none"> • English in the New Zealand Curriculum • Assessment Policy to Practice • TIMSS results published New Zealand performance was below the international mean in Mathematics and slightly above the international mean in Science. This led to the introduction of the Numeracy Project
1995	<ul style="list-style-type: none"> • Technology in the New Zealand Curriculum • Assessing students achievement: ERO provides the first major critique of assessment of the NZ Curriculum • National Education Monitoring Project begins • Social Studies in the New Zealand Curriculum
1996	<ul style="list-style-type: none"> • First assessment resource bank items made available to schools
1997	<ul style="list-style-type: none"> • School Entry Assessment
1998	<ul style="list-style-type: none"> • Assessing Children's curriculum achievement: ERO argues for national testing • Assessment for Success in Primary Schools: National Government's discussion document on national testing • Black and Wiliam publish Inside the Black Box: leading to a subsequent huge emphasis on formative assessment.
1999	<ul style="list-style-type: none"> • Health and Physical Education in New Zealand Curriculum • Report of the Literacy Task Force: too much assessment is effecting teaching • Revision of the NAGs: Response to literacy task force – concentrate on literacy and numeracy in years one to four • Information for Better Learning: Plans for national testing: pilot of externally referenced tests in literacy and numeracy NEMP and ARBs to be expanded • Change of Government: National tests to be developed but their use to be optional
2000	<ul style="list-style-type: none"> • The Arts in the New Zealand Curriculum • Report to the Ministry of Education: A review of the roles and responsibilities of the Education Review Office ERO required to advise and assist as well as review • Count me in too, later the Numeracy Project, begins to roll out in New Zealand schools • The National Assessment Strategy released
2001	<ul style="list-style-type: none"> • Education Standards Act: target setting, planning and reporting • Supplementary Test of Reading Achievement (STAR) released
2003	<ul style="list-style-type: none"> • asTTle released • The Exemplars released
2006	<ul style="list-style-type: none"> • Draft revised curriculum released • Revised PATs

The New Zealand education reforms began with the establishment of BOTs in 1989 (Lange 1988) and were further developed by curriculum reform in the form of

MinZC (MoE, 2002), NZCF (MoE, 2003a) and the NAGs (MoE, 2003c). These reforms were based on a philosophy of increased accountability for schools (Elley 1993, 2004; Dixon, 1999; Gordon & Whitney 1997; OECD, 1983, 1995; Sullivan, 2002; Thrupp, 2005; Willis, 1992) and used a structure known as OBE (Hall, 2005).

OBE required a clearly stated set of outputs for the achievement of schools to be measured against. These outputs were documented in the levels and achievement objectives of the curriculum statements and the contractual arrangement between schools and the government in the form of the NAGs (MoE, 2003). In this sense, OBE and the associated documentation that supported the approach was the major influence on education and, in particular assessment practice in New Zealand, since 1990.

This change in approach to education was adopted widely in many western countries. As has been stated earlier in this thesis, several authors (Elley 1993; 2004; Gordon & Whitney 1997; Sullivan, 2002; Thrupp, 2005; Willis, 1992) consider the New Zealand reforms in education trace back to the UK 1988 Education Reform Act. However, as has also been stated, the New Zealand education reforms differed from those in the UK in two significant ways. Firstly the New Zealand curriculum was implemented without tools to assess learning against the new curriculum or a prescribed system of assessment. This point is covered extensively in Chapter Three.

The second point of difference was that the New Zealand system did not differentiate between what might be considered the “core curriculum,” namely numeracy and literacy, and other learning areas. This distinction was made in the UK and later in the US (Sleeter, 2007). However, in both the UK and the US science was included with numeracy and literacy as forming the core curriculum. The lack of assessment tools and the requirement to treat all learning areas equally were both major influences and challenges during the early years following the reforms.

A further major challenge facing schools in trying to establish assessment systems in the early nineties was the structure and language of the eight levels of the curriculum and the numerous achievement objectives. This point has also been covered extensively earlier in the thesis. To briefly restate this point: the requirement of schools to assess against the achievement objectives in all areas of the curriculum led many schools in the mid-1990s to develop systems that tracked individual children’s achievement against achievement objectives. The format for such systems usually

consisted of some form of three stage marking of the achievement objectives such as: *taught*, *partial mastery* and *mastery* (Baker 1997; Baker & Lorrigan, 1995, 1996; McMaster 1998). The establishment of such systems was often initiated by advisers who provided professional development programmes to support schools in their implementation of the various curriculum documents. The aim of such systems often seemed to involve placating ERO rather than informing teaching. At the turn of the millennium this approach attracted criticism with Hill's (2000a) *Dot, slash, tick: how assessment can drive teachers to ticking instead of teaching* being the most memorable.

As stated above, the lack of a prescribed external assessment system in the New Zealand reforms led to ERO having a great deal of influence on school practice (Dixon, 1999; Fancy, 2004a; French, 2000; Knight, 2000). Dixon expressed strongly the influence that ERO's position had on assessment practices when she stated:

However, it was the Education Review Office (ERO) that was considered to be the agency which enforced these increased demands and requirements. Writers such as Clark (1998), O'Neil (1998) and Robertson et al. (1997) have argued that ERO in its monopoly position as the ultimate arbiter of acceptable practice, have been a driving force in bringing about changes in school and teacher practice, particularly in the area of assessment. (p. 42).

In the late nineties, ERO began a campaign for national testing (ERO, 1998). This campaign was successful in convincing the then National government of the benefits of such a regime which they outlined in *Information for better learning* (MoE, 1999c). This document included a commitment to the development of Exemplars of performance for all areas of the curriculum, the expansion of the ARBs and NEMP, and most importantly a commitment to the development of national tests in literacy and numeracy. While the change of government in 1999 resulted in the use of the tests (asTTle) becoming optional, these developments and the publication of the STAR reading tests (Elley, 2001) meant that after 2000 New Zealand schools finally had a range of assessment tools to measure the curriculum published seven years earlier.

In addition to describing the influence OBE had on the structure of the reforms and the challenges associated with their implementation, Chapter Three also identified four other influences on assessment practice in New Zealand that stemmed largely from international movements. These include authentic or performance assessment, formative assessment, evidence-based-teaching, and international and national testing.

During the years that followed the curriculum reforms, a number of schools began collecting students' work in folders. These collections range from loose collection of students' work and assessment results to using portfolios as described by Arter and Spandels (cited in Ward & Murray-Ward, 1999):

A purposeful collection of student work that tells the story of the student's efforts, progress and achievement in given areas. This collection must include student participation in the selection of portfolio content; the guidelines for selection; the criteria for judging merit, and evidence of student self-reflection. (p. 193)

Individual folders of student work and assessment data were used by schools to demonstrate to ERO that they were meeting the NAG requirement to maintain individual records.

The approach to teaching termed formative assessment has been an extremely strong influence in New Zealand education since 2000. It has had strong endorsement from government education agencies (ERO, 2007; MoE, 2007b) and has been the major content of LPDP and ATOL professional development contracts.

A major influence from international testing was New Zealand's poor TIMSS results in 1994; these results influenced the development and introduction of the Numeracy Project (Higgins, 2002), which has become a major component of primary school assessment.

7.3 Influences and Challenges in the Three Case Study Schools

While the above discussion of the influences and challenges associated with assessment reform has been derived from an analysis of the assessment literature, it is worthwhile considering how these issues have been handled within the three case study schools. Data gathered in the schools tends to focus on relatively recent developments in assessment.

7.3.1 Structure of the Curriculum/Achievement Objectives

As stated above, the requirement of schools to assess against the achievement objectives in all areas of the curriculum led many schools in the mid-1990s to develop systems that tracked individual children's achievement against achievement

objectives. At Endeavour, where the process is still used in school-wide collection of assessment material in the non-core curriculum areas, students' achievement is categorised as *taught*, *consolidated* and *mastered*. This practice may be less common in schools these days and was discarded at Tasman in the recent reform of its assessment system. At Kupe there has been a movement from this three stage form of assessment against achievement objectives to assessing against the levels of the curriculum using the terms developed by asTTle: *three basic*, *three proficient*, *three advanced*, *four basic* etc.

7.3.2 Portfolios and Student Folders

All three schools operated a system of portfolios or student folders which comprise samples of student work, most often self-selected. While portfolios are used in all areas of the curriculum their use is often directed towards assessment in the non-core curriculum areas. At Kupe, student portfolios were the most pervasive form of assessment and received the largest amount of discussion in the interviews with classroom teachers. The school had recently moved from a system where a sample of students' work was collected from all seven curriculum areas each term to a process whereby teachers and children choose the most appropriate sample. Two teachers estimated that this meant they were now collecting between six and eight samples per term for each child. On each sample the teacher made an annotation describing the work and the success criteria.

Linn and Gronlund (2005) and Ward and Murray-Ward (1999) state that portfolios encourage self-assessment and reflection on practice by students as well as being an effective means of communicating with parents. Richard (Endeavour) reported that the introduction of what they termed "Quality Books" had improved children's ability to self-assess. The view that portfolios develop children's ability to self-assess is also held by the two assessment leaders at Kupe and by Judith who teaches new entrants at Tasman.

7.3.3 The Assessment of the Non-Core Curriculum

The three case studies illustrated that assessment and reporting in areas other than literacy and numeracy raised questions of manageability and utility that stem from the original reforms.

The 2005 ERO Report at Tasman School included investigation of the Health and Physical Curriculum as an area of government interest. The review team's comments regarding assessment in this area in many ways illustrate the Office's views on assessment in the non-core areas. The ERO report recommended a need for the school to be assessing students using procedures such as NEMP and the Exemplars. As stated in the Tasman case study, the implication from this comment is that schools should be formally assessing all seven areas of the New Zealand Curriculum on an annual basis. Looking beyond Tasman, if this expectation is taken seriously in schools it runs the risk that data collected has little utility. Furthermore, this expectation that comprehensive assessment data be collected in all areas of the curriculum annually, presents a further challenge for schools in terms of the utility of what is recorded in that the common approach to timetabling and the teaching of these subjects is through blocks of five to six weeks under the heading *Topic*. For example, a cohort of year 4 students may have been taught a science unit on "*Planet Earth and Beyond*" in May 2006, but they would be unlikely to encounter a unit in that same strand of the science curriculum until 2008 when they are in year 6. Achievement data collected at the end of such blocks of teaching may be used to report to parents, the BOT and for ERO review purposes. However, because the students have now moved on to another area of learning and will not revisit this content for another two years the information gathered can contribute little to teaching in any formative manner. Continuity in assessing achievement and progress is therefore absent and has little, if any impact on teaching. This point was made by the assessment leaders at Endeavour when they dismissed advice regarding assessment of Social Studies that they received at a recent professional development course.

The theme of the lack of utility of assessment in non-core areas was raised by a number of teachers. Mary from Endeavour comments on teaching science and social studies.

Talking before the science and social studies it is not the thing that is particularly meaningful. It is not going to help the child and it is not going to help me as a teacher to improve my practices necessarily because I am not going to revisit the topic after I have assessed it.

It again illustrates that there is a gap between the ideal and the reality in assessment.

7.4 Three Waves of Assessment Practice

The framework entitled the three waves of assessment practice (Figure 7.1) lists the influences on, and challenges associated with, the implementation of assessment practice in New Zealand schools that formed the first research question.

This framework was developed from the review of the literature on assessment in New Zealand schools and the investigation in the three schools. The framework was subsequently examined by the three focus groups of 15 experienced principals who had taught during this period. Where members of the focus groups differ from or add to the original framework these views are recorded.

First wave: 1989-1996

From the passing of the Education Act in 1989 to 1993 schools were mainly focused on the establishment of Boards of Trustees and the writing of charters and policies. The focus on the curriculum and assessment did not become intensive until 1992-1993 with the publication of the first curriculum statement (Mathematics, 1992) and the NZCF (1993), and the promulgation of the NEGS and NAGs (1993). Faced with this new demand for assessment and reporting to their BOTs, schools fell back on those assessment systems they were familiar with (French, 2000; ERO, 2001). This included PAT tests, Running Records, formal spelling tests like Daniels and Diack (1958) and Peters (1979), and the schools' own basic facts tests. While this period saw an increasing amount of assessment being undertaken, a lot of it was very piecemeal and different areas or syndicates of the same school were carrying out different processes or using tools quite differently. Professional development contracts that were available to assist schools to implement the new curriculum suggested the development of the triangular type tick sheets spoken about earlier (Hill 2000, 2000a). Their use in mathematics especially was very common. Overall schools were confused and frustrated. There was confusion over how to assess in areas that had been subject to little formal scrutiny before and no assessment tools were available for the new curriculum statements that schools were being asked to implement in rapid succession (Fancy, 2004a). The research of Livingstone (1994), Renwick and Gray (1995) and Wylie (1997 1999) captures the frustration of those teachers working in schools during this period.

Members of the focus groups also noted that the period from 1989 to 1993/4 was mainly concerned with the development of charters, and policies and getting to grip with financial and property self management. From 1993 on schools were grappling with the new curriculum requirements which included large elements of assessment. Assessment also became a focus as more schools had ERO reviews and the word spread among the principal community that there were high expectations about having data to show them. Members of the group also commented that for some schools the practice of establishing individual student folders and achievement objective tracking sheets began several years before 1996. They felt that while the stages captured what had happened, the division between the first and second wave was perhaps more fluid and that the period when a number of schools had moved to the second wave was in 1994-1995. However, they also stated that a smaller number of schools were later than 1996 in making similar changes.

Second Wave: 1996-2002/2003

The second wave saw the establishment, in most schools, of a more systematic approach to data collection with schools moving to school-wide systems. It was during this period that a number of schools formalised their tracking of individual children through folders of student work and test results that accompanied the child as they moved through school. The amount of assessment material collected greatly increased as schools responded to ERO's push for a balanced curriculum, demonstrating that schools were assessing against all levels of the curriculum statements, and that BOTs were making a difference to student achievement.

Third Wave: 2002 – 2007

The third wave saw a massive change in assessment systems and processes in primary schools. This included the use of newly developed normative assessment tools such as STAR and asTTle as well as the implementation of the concept of formative assessment. These two developments were supported by a large number of Literacy and ATOL professional development contracts. The wide adoption of asTTle Writing introduced teachers to the concepts of moderation and the need for systematic collection of data. By 2007 the Numeracy Project and its associated assessment had been introduced to almost all New Zealand schools. In this period ERO frequently recommended to schools to move from merely collecting and reporting achievement data to analysing the results in depth. Schools, individual teachers and management

increasingly question the purpose of retaining past assessment practices. However, within the three case study schools there seems to have been some reluctance to abandon some of the tools and practices of last century.

While some members of the focus group questioned the appropriate dates for the transition from the first wave to the second wave these principals were very much in agreement with the period around 2002 being the start of the third wave. They described the use of STAR, asTTle and the use of the concepts and language of formative assessment as marking the change.

The focus groups agreed with the idea that during the third wave teachers and in particular school leaders felt more confident and knowledgeable about assessment. They recorded that they felt their practices in this area were now more based on research. Such views support the conclusions of Hipkins, Joyce and Wylie (2007).

Figure 7.1: Three Waves of Assessment Practice in New Zealand Primary Schools 1989-2006

<p style="text-align: center;">First Wave: 1989-1996</p> <ul style="list-style-type: none"> ❖ 1989 Tomorrows Schools - Formation of ERO, First ERO Reports - school asked to show they are teaching and assessing a balanced curriculum and making a difference to students' learning. ❖ 1992 Mathematics in the New Zealand Curriculum is introduced. ❖ 1993 New Zealand Curriculum Framework, NAGs & NEGS introduced and gazetted. <p>Common School Responses:</p> <ul style="list-style-type: none"> • Increased use of Running Records, spelling tests, PAT Tests. • Start of individual folders and increased assessment in non-core curriculum. • Collection of assessment data, although not often systematic throughout the school. • Little analysis of assessment data. <p><i>School staff report frustration with assessment expectations (Renwick & Gray, 1999; Wylie, 1997, 1999)</i></p>
<p style="text-align: center;">Second Wave: 1996-2002</p> <ul style="list-style-type: none"> ❖ 1999 Literacy Task Force raises concerns about high levels of assessment in the non-core curriculum. ❖ NAGs revised – 1999. ❖ Numeracy Project begins in selected schools – 1999. ❖ Planning and Reporting requirements – Education Standards Act 2001. <p>Common School Responses:</p> <ul style="list-style-type: none"> • Schools moving to a more systematic approach to data collection with the development of school-wide systems. • Schools formalised their tracking of individual children through folders of student work and tests results that accompanied the child as they moved through school. • The amount of assessment material collected greatly increased as schools responded to ERO's push for a balanced curriculum, demonstrating that schools were assessing against all levels of the curriculum, and that BOTs were making a difference to student achievement. <p><i>School staff report frustration with assessment procedures (Wylie 1997, 1999)</i></p>
<p style="text-align: center;">Third Wave 2002/2003 -2006</p> <p>Development and introduction of STAR, asTTle and The Exemplars; Numeracy Project continues and coverage extends to most schools by 2006</p> <p>Common School Responses:</p> <ul style="list-style-type: none"> • Large emphasis on formative assessment • Planning and reporting - target setting implemented • Increased assessment and analysis of data • Widespread use of asTTle and STAR • Movement to more electronic recording of assessment data • Numeracy Assessment procedures increasingly common <p><i>School staff more positive regarding assessment procedures (Hipkins, Joyce & Wylie, 2007)</i></p>

Research Question Two

How did three New Zealand primary schools establish assessment and reporting systems during the period 2003-2007 that were manageable and maintained a focus on learning?

7.5 Factors that Influence the Establishment of Assessment and Reporting Systems

The research identified five major factors that influenced the establishment and modification of assessment practice in the three schools.

7.5.1 School Leaders

The philosophy and experience of the schools' leaders was a major influence on the establishment of assessment systems and practices in the three schools investigated in this study. This influence is particularly evident in the way assessment systems were developed at Kupe and Endeavour schools. Rebecca, the principal of Kupe School and Richard from Endeavour have more in common with each other than with Brian at Tasman. Both Rebecca and Richard held senior but not principal positions prior to and during the transition to Tomorrows Schools. They were appointed to their present schools in 1996 (Richard) and 1997 (Rebecca). Both reported problems with assessment in their respective schools on their arrival. The tools being used were similar in both schools; Running Records, spelling tests and school developed mathematics basic facts tests. Rebecca and Richard both believed these assessments were being used ineffectively and Richard's view was that there was not enough assessment being undertaken. Both Richard and Rebecca were supported in developing assessment programmes by long serving and experienced deputy principals.

Richard and Rebecca can both be represented as establishers of assessment systems. A pragmatic approach meant that both established assessment systems that largely worked for their schools. An examination of their school's ERO Reports and their interviews indicate that the schools were successful in meeting their legislative requirements.

While there were many similarities in their approach and the assessment systems they developed, the journey each school took differed in several ways. Rebecca began her

tenure at Kupe by engaging in a school-wide focus on a shared understanding of teaching and learning. From this exercise the school philosophy and procedures on assessment were developed. As Rebecca noted:

So we went right back to the basics. What is effective learning and teaching? What does it mean for you and you and you? We came out with some principles that underpinned what we did. So we said we have got to know where kids are at so how do we find that out? What tools do we use?

Richard's approach to assessment and change was more measured and was evolutionary rather than a dramatic change. During his 10 years at Endeavour Richard established a comprehensive assessment system that met the needs of his particular community. Richard described assessment when he arrived at Endeavour in 1996 and progress since then in the following way:

There was no school-wide assessment so I started to put those sorts of things in and strengthen practices within the school so they were much more consistent around the school, rather than individual teachers doing different things which were happening before I came. A lot of schools were like that I am sure. It was more of an evolving process rather than a huge deliberate focus. That's the way it tended to be, there was no whole year focusing just on assessment.

In contrast to Rebecca and Richard, Brian faced multiple challenges when he arrived at Tasman. The school had a challenging ERO report, and there was almost a complete change of staff including both the deputy and assistant principal. In addition to these challenges, neither Brian nor his newly appointed management staff had the length of teaching and management experience in 2005 that Rebecca and Richard were able to call on nine years earlier. It is important to note that changes in the assessment framework at Tasman were being made in a far shorter time-frame and were more radical in nature than what happened in the other two schools in the period 1997-2000. A close reading of Tasman's ERO reports since 1993 shows that the school had experienced high staff turnover and included comments that assessment procedures often fell short of what ERO expected.

It is worth noting that the influence of leadership on assessment practice is not solely confined to the principal. In the postscript interview with Brian, the principal of Tasman, noted that he, the deputy and assistant principal had been together for nearly five years now and that contributed a lot to a feeling of greatly increased confidence about assessment decisions and programmes in the school.

The focus groups strongly supported the view that beliefs and understanding of the principal had a major influence on assessment practice in the school. Several members of these groups stated that this was particularly evident when the school experiences a change in principal. The appointment of a new principal, in their experience usually led to a change in direction for assessment in the school.

While Rebecca and Richard can be seen as school leaders who established assessment systems in their schools in the late 1990s, the case studies identified a new group of teachers who had leadership roles but differed in their perspective on assessment and their experience in teaching. Most of these people had not begun teaching at the time of the reforms. They had rapidly risen to middle and senior management positions and were often leaders in implementing electronic recording of assessment. This group included Ross (AP, Tasman), Judith (DP, Tasman), Maria (syndicate leader, Kupe) and Susan and Mary (assessment leaders, Endeavour). This group was frequently concerned with the utility of assessment which is exhibited in the following quotes from the three cases studies regarding the purpose of assessment.

We use Peters (spelling test) at the moment and one of the questions is what does it show so why do it? We don't do the analysis properly. We just take the test and record an age. The teachers are saying well that doesn't tell me anything where they use blends or how they are putting words together so why are you doing it? Is it just to get an age but that doesn't help me with my teaching. (Judith).

That seems to be the change in thinking. Now if you can't tell me why we are doing it –well then we are not doing it. (Ross).

So Susan and I have identified changes in our current assessment procedures that we think are necessary. The main one is the cumulative cards and how we use those, what we use those for and we are also implementing classroom manager. Why have another piece of paper that effectively nobody looks at and that goes back to what is its purpose why are we using it, how is it going to inform what is going to happen. (Mary).

Maria described the challenge of getting some staff members to think about assessment:

From my perspective from someone who has gone from being a scale A to management it can be the personalities of a teacher who will not do something unless they are told they have to do it. If they are left with a choice they won't know what to and others go I want to make my own choice and so it like trying to balance those needs and there is a conflict that happens. To be quite honest, it's those people who have been around a bit longer that have been harder to shift because they do want to be told what to do. They don't

want to shift out of their normal triangle way of assessing, and using levelling, and all the rest of it because they haven't quite got their heads around it.

These new school leaders are often the teachers in the school who will experiment with, and initiate the use of new assessment tools such as the ARBs, NEMP tasks, and the Exemplars in their classrooms. While this group did challenge the utility of assessment, overall they were enthusiastic about where their schools were going with assessment. While there were debates with their principals on the utility of some of the assessment tools and practices, such debates could not be described as conflict.

As stated earlier, the original design of this research did call for more scale A teachers to be interviewed. Problems with staff turnover at Tasman and the decision to interview the assessment leaders at Kupe rather than scale A teachers meant that the research did not get a strong perspective on the views of basic scale teachers. The two scale A teachers at Kupe and the deputy principal at Endeavour, while not wholly negative on assessment, shared the same concerns as the new leaders on the utility of some assessment tools but were generally not as optimistic about assessment, nor were they as likely to be aware in detail of other tools and they were less likely to experiment with new tools in their classroom.

7.5.2 Professional Development

All three schools were influenced strongly in their assessment practice by recent professional development contracts. The three schools were participating, or had recently participated in major school-based professional development contracts that had introduced new assessment tools and influenced assessment procedures: Tasman (guided reading LPDP contract, asTTle Reading), Endeavour (writing, LPDP contract, asTTle Writing and the Exemplars: Writing), and Kupe (Numeracy Project, ATOL, asTTle Writing). All three schools were enthusiastic about the benefits of this professional development and the associated development in assessment. The three principals noted how formative assessment, which was introduced in the professional development contracts in specific areas of the curriculum, was now being applied to all teaching areas in the school. This was most notable at Endeavour and Kupe. For Endeavour all those interviewed felt that the school had attained a lot of understanding out of the recently introduced writing assessment. From a manageability perspective they acknowledged that the assessment and moderating

procedures were very time-consuming, but in terms of utility they reported that the quality of the information and the understanding gained was worth the considerable time they devoted to such activities. Richard contrasted this professional development with what he considered was a less successful structure with the Numeracy Project, from which he felt he had not developed a comprehensive understanding of assessment in this programme. This caused him to question the utility of the information gathered. At Tasman, Brian felt that the LPDP Contract and the subsequent implementation of asTTle Reading had enabled the school to gain more useful assessment information and made the staff more sensitive to the assessment of children for whom English was a second language.

The focus groups of principals stressed that there was a strong connection between successful professional development programmes and effective implementation of changes in assessment practice.

Participants' Assessment Background.

The interview process asked all participants where they had gained their knowledge of assessment. As noted above, most participants indicated the strong influence of recent professional development contracts in contributing to their understanding of assessment. Other than the contracts it was common for teachers and principals to report that they had acquired their knowledge in an on-going, and often informal manner. In terms of scale A teachers Jill from Kupe gives a response that was evident in the interviews of some other participants:

I don't remember doing any assessment in my first few years of teaching. So when I came back into teaching people assumed that you knew and I didn't. So I learnt on the run with lots of listening and a couple of staff meetings where we went over Running Records. I have just learnt as I have gone on teaching and I think probably I am still in the process of doing that. My understanding of Maths assessment has probably developed as part of doing the Numeracy Project when I did writing it came from the Literacy Contract.

Richard, the principal of Endeavour describes how he has developed his understanding of assessment in an informal and on-going manner.

Basically what I have learnt about assessment has been on the job, doing it. As a classroom teacher coming through and learning from other people. I have been to some courses that had a component of assessment. I don't think as I look back on it I went to any courses purely on assessment, a lot of meetings readings etc. In terms of readings I have been influenced by what the Ministry sends you, I have been trying to think what sticks out. There

have been a lot of things that have been read and amalgamated into my way of thinking. So there has been no sort of particular guru that I have followed.

I have been influenced by talking about assessment with other principals, teachers at other schools talking about what they have been doing. A sort of amalgam of influences, not one particular school or person. There has been some the other way around; other people coming to us and hearing what we do.

Rebecca, the principal at Kupe, had a more structured background and mentioned ASTU papers she had done on assessment and her time as an advisor. Several of the participants mentioned the influence of such books as Clark, Timperley and Hattie's (2003) *Unlocking Formative Assessment* and recent courses led by Auckland educator Mai Lai.

The responses to questions on how the participants had gained their knowledge on assessment did not lead to direct statements on manageability. However, there was an indication that teachers and principals had become more sensitive to the utility of assessment material as evidenced by increased questioning of the purpose of assessment in all three schools.

In the focus groups the principals mentioned the strong influence of a number of individuals, often professional advisers on assessment practise during the nineties. They concluded that principals and teachers were now better informed on assessment issues and less likely to follow the advice of a single person.

7.5.3 Networks with Other Schools

Formal Networks

While competition between schools was seen initially as a key element of the reforms (Elley, 1993, 2004; Fancy, 2004a) it is now common for schools to be encouraged by the MoE to work together in the area of assessment and professional development (Fancy, 2004b). Kupe School was involved in a cluster of eight local schools that undertook analysis of comparative data from STAR results through a Ministry of Education initiative. As a result of this cross school data gathering, the principal of Kupe was more sensitive to issues of validity in assessments and was the only respondent who used the term *clean data* when talking about assessment results:

I think having everybody being clear about administration especially those norm-referenced tools and building that common language and making sure the data is clean.

Endeavour School joined three other local schools in adopting the “Quality Schools” programme which resulted in the use of Quality Workbooks for assessment and reporting.

The major influence that formal networks had on changing assessment practice in participating school was stressed by several members of the focus groups who had been involved in projects with schools in their cluster.

Informal Networks

Almost all participants mentioned the influence of informal contacts with other schools, teachers and principals as being of some influence on their own school’s assessment procedures. This was illustrated by Richard, who on the day of the interview, reported that he had just been at a principals’ gathering that morning and that a discussion with two other principals was leading him to consider increasing the number of administrations of STAR that the school did each year.

7.5.4 The Influence of ERO

The Education Review Office has been a powerful influence in the development of assessment in New Zealand schools. This is illustrated from an analysis of all the available ERO reports for the three schools. As noted in previous chapters, Tasman, received the most challenging ERO reports. At the time of the data collection for this research, the school was responding to the 2005 report which included a major recommendation about assessment. Their response included the implementation of asTTle Reading and Mathematics as well as STAR as a twice yearly activity; this response could be considered to be an overreaction to the ERO report. Endeavour, which has had positive ERO reports with some recommendations on assessment, has reacted to ERO’s views in the measured way that characterises this school’s approach to assessment in general. Perhaps not surprisingly, Kupe, which undertook more assessment than the other two schools, had received a succession of glowing reports regarding its assessment procedures.

7.5.5 Ministry of Education

As stated a number of times in this thesis, a defining characteristic of assessment in New Zealand primary schools is the degree of openness in assessment choices that a school can make. In comparison with assessment in the United Kingdom, which was the model for many of the Tomorrow's Schools reforms, New Zealand schools have a great deal of choice on how they assess and report student achievement. The Ministry of Education does not directly tell schools which assessment tools to use. However, they do devote considerable resources to publicising new assessment tools and training schools in the use of new assessment tools. As stated above in the section on formal networks with other schools, a significant feature in the last eight years has been the Ministry of Education supporting clusters of schools to work together in an evidence-based teaching approach. Fancy (2004a) described this as schools moving from being isolated islands to interdependence. This is evident in a number of government initiatives such as the *Picking up the Pace* programme in South Auckland (Phillips, McNaughton & MacDonald, 2001), and both the I CAN (Porirua) and the Taita Central Literacy Project in Wellington (Mallard, 2000, 2003). As stated above, Kupe is involved in an initiative similar to these.

In 2003 the Ministry of Education introduced the "Planning and Reporting" process which involved annual setting of targets. These were checked by the local Ministry of Education Office, as was the resulting statement regarding the achievement of those targets in each school's annual report. This has been the only direct compliance requirement for individual schools. It must be said that even this process allows schools a lot of individual choice and variety of models (Hipkins, Joyce, & Wylie, 2007; Thrupp, 2004).

Brian (Tasman) illustrated the Ministry's role both directly and indirectly when he first came across annual targets after spending three years overseas:

The targets were a bit of a shock to the system in 2003. I got some very good PD as a new principal from a person in the Ministry when I was down in (South Island town). She basically said I haven't got the targets. Because I had been away for three years and I had missed all that, those red flyers that came out. I said what the hell are you talking about and didn't realise what

they were and she came down and spent a day with me in January 2004, in the holidays and from then on it has been fine.

Rebecca (Kupe) spoke of how the Ministry sponsored electronic website *Leadspace*, which she reported visiting about twice a week, had developed her understanding of assessment.

The focus groups commented that the MoE increased involvement in assessment particularly with the promotion of particular assessment tools was a defining characteristic of the third wave of assessment change.

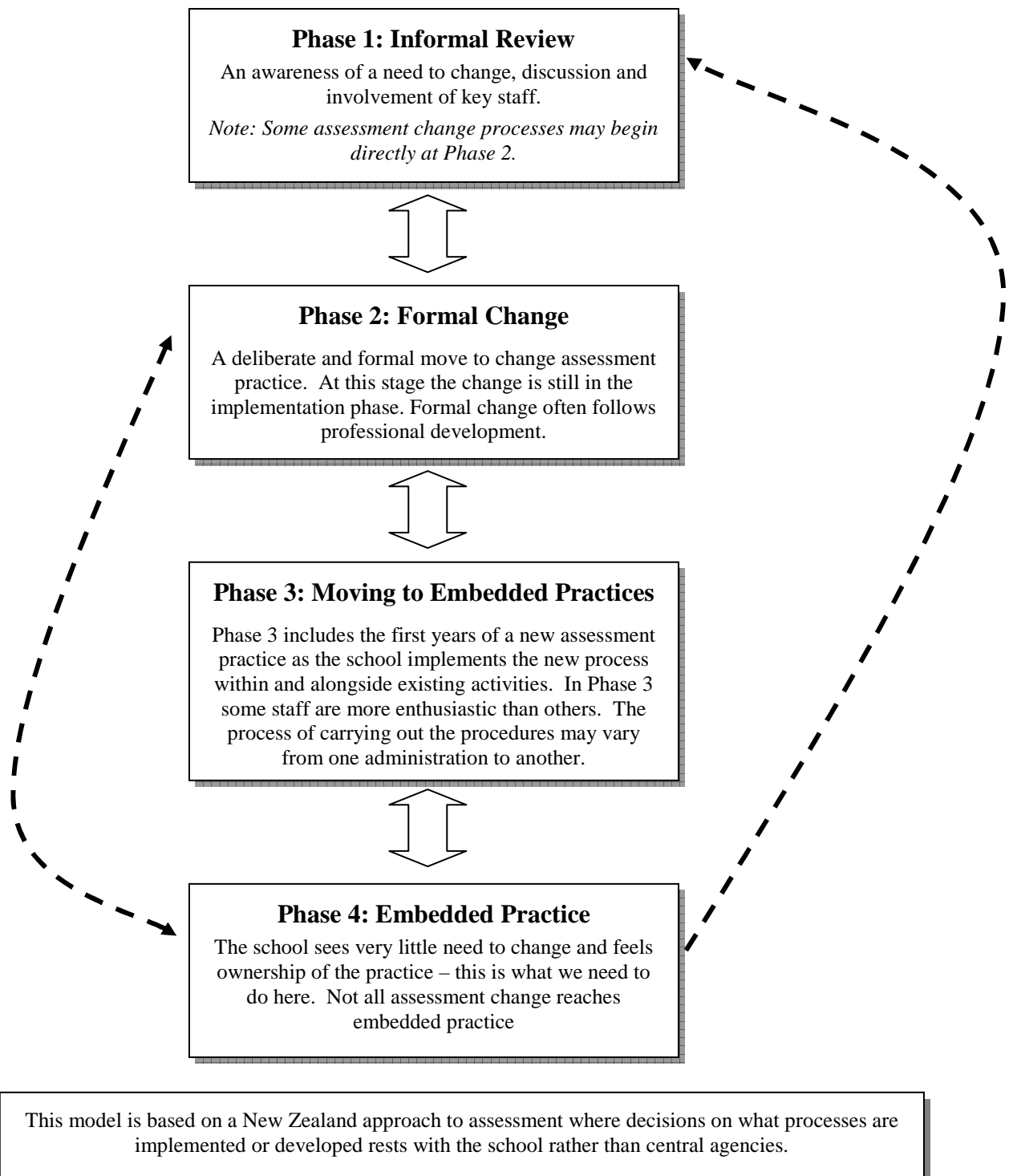
7.6 Four Phases of Assessment Change

The framework entitled *Four Phases of Assessment Change* (see Figure 7.2) uses interview data from the three case study schools to illustrate how assessment and reporting systems appear to have been established and modified. It should be noted that ongoing change appeared to be characteristic of all three schools. The original framework was developed from the investigation in the three case study schools in the period 2007-2008. During 2009 the three focus groups reviewed the four phases framework and their members' views are noted at the end of this section.

Four stages seem evident:

1. Informal Review – an awareness of a need to change, discussion and involvement of key players.
2. Formal Change which has not moved to embedded practice – a deliberate formal move to change assessment practice, but the change is still in the implementation phase.
3. Moving to embedded practice.
4. Embedded practice – the school sees very little need to change and feels ownership of the practice: *this is what we do here*.

Figure 7.2: Four Phases of Assessment Change in New Zealand Primary Schools



A theme of this research is that schools, like all organisations, tend to be complex and not always systematic in their implementation of new practices. Assessment activities in the schools were fluid. Different aspects of the schools' assessment procedures and areas of the curriculum were at different stages. For example, while the assessment of writing in two of the three schools could be described as at stage three – moving to embedded practice - all three schools were in either stage one or two in assessing mathematics.

While the above model has a four stage process it would also be misleading to think of the process as being always linear. In one school there was a clear example of an assessment approach that reached stage two but was then discontinued. In a response to a question on the assessment of higher order thinking. Brian, the principal of Tasman School since 2005, stated:

There was a woman here in 2004 and she was very big on inquiry learning and she really pulled the staff in. When she left and I came here and they told me all about it. I kind of expected them to get on with but it fell by the way side.

Stage One: Informal Review

At this stage there was an awareness of a need to change which may have involved the school's use of a new assessment tool, the use of an existing tool in a different manner or the discontinuing of an existing assessment activity. The interviews gave the impression that a great deal of this came from personal reflection by the principal or other staff, informal contact with other schools, and discussion among the key players such as the deputy and assistant principal.

Richard, the principal of Endeavour School, illustrated this stage in a number of his responses. On Running Records he noted: "In terms of reading we have also been doing Running Records throughout the school but I have been thinking about a better way." In other curriculum areas he observed: "We want to look at something in maths where we feel we are a little light." In relation to the Numeracy Project Assessment he commented: "I have seen the results. The teachers have given to me the sheets they have done. So I think to myself, how can I use this as a school wide thing but I can't get my head around that and work out what to do."

The first stage is also evident in Kupe School, which joined the Numeracy Project in 2006. Rebecca, the principal, reflected on how mathematics will continue to be

assessed in the school: “We used to have basic facts; we have them in this year but we are thinking we might take that out because it is giving the wrong message about what is important.”

Stage Two: Formal Change

At this stage each school made a specific decision to change an important assessment practice but the change was still in the implementation phase. Such a change commonly occurred during and immediately following a professional development contract. At this stage some teachers and/or areas of the school tended to be further advanced than others.

Richard’s response to a question on how frequently teachers were expected to undertake Running Records illustrates a school moving from stage one to stage two:

We used to but now for this year the seniors are experimenting with the asTTle Reading test. We haven’t enforced the running record policy up there and that will change formally, probably next year, as we learn more about asTTle Reading which is going to be much more useful.

Brian’s description of what was happening in his school with the GLOSS testing in the Numeracy Project illustrated this transitional phase: “They did three last year. Twice this year. It has given us enough data to convince us that twice is enough.”

Stage Three: Transitioning to Embedded Practice

At this stage the assessment practice had usually been in operation for more than two years and the school was committed to its continued use. For example, Endeavour had used asTTle writing for three years and while it continued to develop its use further, it could be described as being at stage three.

Grace, the Endeavour Deputy Principal illustrates the concept of moving to embedded practice in this statement about the introduction of asTTle writing:

We are coming to grips with the AsTTle Writing. When we first started it; it was pretty harem sacrum because it was pretty messy. We needed and we got lots of input from the facilitators who were running the literacy programme but it has taken two or three years to get rid of all the teething problems. That’s going well and the asTTle Reading programme we have just started that so we are still going through teething stages with that.

One of the challenges during this period is getting consistent assessment practices from staff. As Mary from Endeavour states: “I have found some of the challenges is

getting consensus from teachers and getting uniformity as well making sure people are doing exactly what they should be doing.”

Stage Four: Embedded Practice

In this phase the assessment practice has clearly been embedded into the on-going operation of the school. Richard described Endeavour’s attitude assessment survey as an approach that had become part of the ongoing practice of the school with a strong sense of ownership that characterises this phase of practice.

We do an annual survey of kids from year 4 to year 8 and a sampling of kids from year 3 about their attitude toward school, their attitude toward the curriculum, bullying, etc. things like that to inform us how they are thinking, what they are thinking. That has been quite good to assess how our major changes in writing have impacted on the kids.

At Kupe, the PAT Listening Comprehension test could be categorised as being at stage four in that it had been carried out for a number of years and has been retained after a number of other PAT tests had been discontinued after a recent review.

So we used to do PAT Listening Comprehension, Reading Comprehension and Vocabulary and Maths in March. Now we do Listening in March because we think it is useful for triangulating information. We don’t do Reading Comprehension and Vocab now. We will be doing Maths for one more year.

7.7 Manageability and Utility in the Case Study Schools

While the above model addresses the establishment and modification of assessment systems in the schools there is a need for this analysis to address the second part of the first research question: “have these recent developments produced systems that are both manageable and maintain a focus on learning?” The answer to this part of the question is mixed. In terms of the focus on learning there were positive elements. The schools in this study that have undertaken professional development in writing are extremely enthusiastic about the assessment process. It has, they have reported, produced information that has a high focus on learning. This would be also be the conclusion of the staff at Tasman school on the utility of the data gained from asTTle Reading. Furthermore, teachers reported that the opportunity to work co-operatively with others has brought a new depth to their teaching. In a similar manner the concepts associated with formative assessment have brought a new energy to many of the subjects being taught and teachers would agree that they have been central to maintaining a focus on learning. A number of teachers, though certainly by no means

all, report that the Numeracy Project Assessment, especially the GLOSS has contributed to an assessment system that maintains a focus on learning.

However, there are many elements present in the three schools' assessment systems that are questionable in terms of manageability and utility. As stated a number of times, much of the assessment of non-core curriculum areas appears to be of questionable value. There is still evidence of replication in the recording of data with schools having two and even three different systems of tracking individual children's progress. While the schools report their portfolio systems are valued by parents, this is a process that consumes a lot of teachers' time yet a number of teachers report that it contributes very little to a focus on learning.

Despite almost all of the teachers and leaders who were interviewed questioning the value of administering Running Records to competent readers, all schools persisted to some extent with this practice. This raises questions in terms of both manageability and utility. Similarly, although all participants reported an enthusiasm for formative assessment, a number seriously questioned the manageability of recording feedback in writing when this had been suggested by recent ERO visits to the school. As Brian from Tasman stated:

I mentioned feedback before, I guess written feedback has its levels of quality. That's something that needs to be very consistent but that is very time consuming and I've tried it. Sat down in a handwriting lesson in a junior class and ERO had been mouthing off to us you have got to write down what the next steps are for the children and I sat down and what use to take me as a teacher 20 minutes a lesson I sat down and wrote then written feedback I had a queue out the door.

It is a concern that the Numeracy Project assessment has left a number of teachers and two schools unclear about an appropriate on-going assessment regime.

While most classroom teachers understand the rationale for the collection of data for accountability purposes they find these assessment activities questionable in terms of time and utility.

On a more general note, participants in this research were also asked if they thought the assessment requirements imposed on their school, and on them personally, were reasonable and manageable. It is difficult to give a definitive answer to this question as it raises emotional issues for teachers such as workload and work-life balance. A lot of those interviewed do find many aspects of assessment unreasonable. In

particular, many of those questioned find the summative reporting that is required for reporting to BOTs and other statutory bodies frustrating and unreasonable.

However, it also needs to be stated that the teachers are more positive and optimistic about assessment than appeared to be the case in Wylie's surveys in the late nineties or in Dixon's research with Auckland teachers (1999), and Renwick and Gray's case studies (1995). This conclusion is supported by recent research which, although focused on Planning and Reporting, did allow Wylie (pers comm. 27 July) to compare teachers' current attitudes to assessment with her survey results eight years earlier (Hipkins, Joyce & Wylie, 2007). As this team stated: "Most primary teachers are confident about their ability to interpret achievement data and there has been a marked increase in self-reported confidence since the 1999 NZCER National Survey (Wylie, 1999) (ii).

While teachers are more positive about assessment than they were 10 years ago, there are a number of manageability and utility issues that remain.

7.8 Manageability and Utility in New Zealand Primary Education

The section above considered the issues of manageability and utility in the context of the three case study schools. To gain a fuller appreciation of these issues, it is helpful to examine these concepts from both the case study data and the wider literature. Figure 7.3 presents these relationships in a visual format which will be discussed below. As well as exploring these relationships Figure 7.3 also attempts to illustrate the relationship between manageability, utility, reliability and various concepts of validity that were initially examined in Chapter Two.

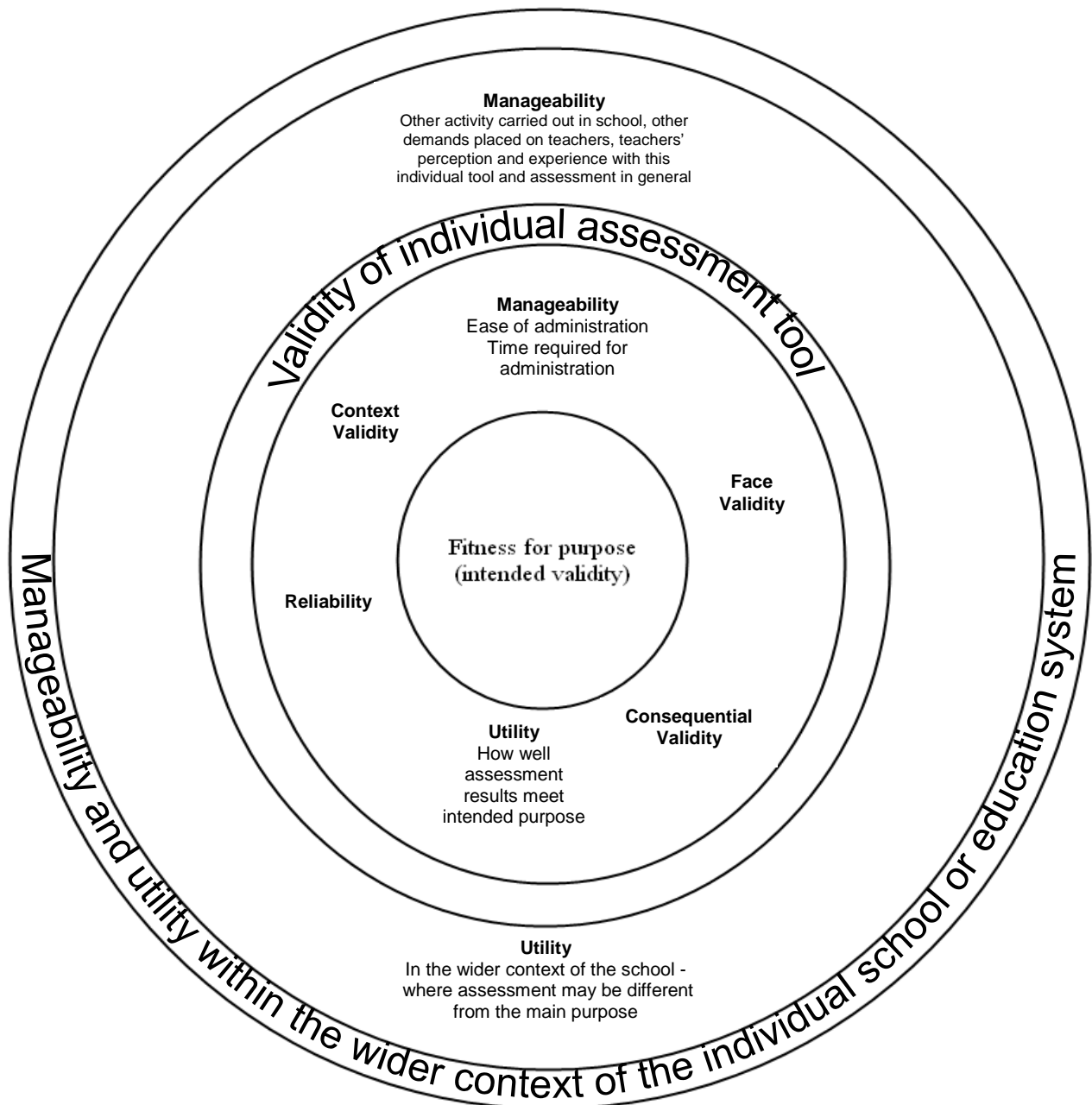
At the core of this is the key notion of validity: fitness for purpose. However, diagram highlights the need to view manageability and utility from two perspectives. In the inner area (under the label of "validity, of individual assessment tools") the two concepts are examined in line with the question: "is an individual test or assessment procedure manageable and/or useful?" The outer area asks a related but different question: "given the other assessments carried out in this school, the other demands on teachers, and the perspectives of teachers individually and collectively, is the assessment activity that the school carries out, or parts of it, manageable and useful?"

It is the examination of each circle and the interaction between them that has been the theme of this thesis.

The diagram also makes the point that concepts of fitness for purpose and manageability and utility are closely related. Mary, one of the assessment leaders at Endeavour, captures the decisions that schools have to make on a regular basis regarding the purpose, utility and manageability of assessment.

We keep reminding Richard [the principal] we can do this assessment task but what is its purpose, why are we doing it and what are we going to do with the information we find out? It is important to sometimes ask teachers are you going to do something with the information you find out because there is no point in doing a task that can take a long time. The most challenging thing for me would be to take that time to look at the results in detail and fairness to the children and sit down and give myself time to think quite clearly through what their next learning step would be.

Figure 7.3 The relationship between manageability, utility, reliability and validity



Further illustrations are evident from both Endeavour and Kupe in their evaluation of asTTle writing as demanding in terms of manageability but of high utility in guiding subsequent teaching and learning. On a more complex note, the time devoted to the collection of portfolios at Kupe was reported as being demanding in terms of manageability but was considered useful for reporting to parents but of lesser value in informing teaching.

As stated above, analysis of the outer circle (manageability and utility in the wider context of the individual school and education systems) raises different questions from the inner circle. At the time of the original interviews, Tasman School was administering four formal reading tests in a year: two STAR and two asTTle. In the 2009 postscript interview with Richard, the principal of Endeavour, he reported that his school was similarly using the same tests and schedule for assessing reading. When examined as individual tools (inner circle) STAR and asTTle Reading could both be considered reasonably manageable. However, when the administration of four formal reading tests during a period of 40 weeks of instruction is considered, then the manageability of reading assessment at these two schools is likely to be a challenge. This illustrates the point that in order to gain a perspective on the manageability of assessment in a school, it is necessary to examine not only the use of individual procedures but how much other assessment is carried out.

The above point is also illustrated by the way that the two assessment leaders at Endeavour questioned the necessity of continuing to collect individual assessment information through the use of cumulative cards when the school was collecting this information through other methods.

While the issues regarding multiple reading assessments and schools duplicating systems for maintaining individual records are decisions made at school level, concerns regarding the utility of assessment data gathered in the non-core curriculum illustrate the impact of the requirements of the national education system on each school's assessment procedures. The requirement that schools maintain individual records that contain student achievement information in all learning areas, as set out in the NEGS (MoE, 1993c), led to schools collecting information that a number of teachers in the study questioned in terms of the utility of the information obtained.

In terms of the utility of assessment in the wider context of the school system, it is worthwhile to return to another issue discussed in Chapter Two. There are elements of the utility of assessment that extend beyond how assessment is described. To illustrate this point, assessment procedures such as PAT Tests and Running Records have high utility when used for their intended purposes. However, when these assessment tools were used in the early and mid-nineties by principals to aggregate and report student achievement in relation to the new curriculum, they were often being used for a

purpose for which they were not necessarily well suited. Their utility for such reporting was limited.

As was stated in Chapter Two, when we consider manageability and utility of assessment we are also dealing with teachers' perceptions about these issues. This was illustrated in the literature by Gilmore and Hattie (2000, 2001) when they examined the use of the ARBs and reported that those teachers who made little or no use of the resource reported that it was too time consuming, whereas high users made no comment on this factor. Also of relevance, Brown (2002) reported that some teachers who experience assessment regimes that they perceive to be unmanageable and of little utility may then move to questioning the utility of all assessment practices. As already noted, judgements of manageability and utility involve subjective evaluations on the part of teachers and schools. It is not always a case of how much assessment a school carries out, or how useful it is, but how those involved in these activities perceive them.

7.9 Recent Developments – Research Question Three

To what extent do the staff of three case study schools find the most recent changes in assessment practice/tools manageable and provide useful information?

7.9.1 Formative Assessment

The pervasive use of the language of formative assessment was a characteristic of all three schools. The use of such terms as *learning intentions*, *success criteria*, *next step and feedback* occurred very frequently in interviews. As mentioned above, the schools could trace much of this practice back to specific professional development contracts. In addition, a number of the teachers interviewed could mention specific books such as Clark, Timperley and Hattie's (2003) *Unlocking Formative Assessment* and other titles as influencing their practice in this area. This indicates the value of texts that teachers find accessible in the presentation of ideas.

On a less positive note, two participants highlighted a gap between the ideal and reality in formative assessment. In their *Process Indicators* and their Review Reports, ERO declare an expectation that feedback comments should be written. Two of the teachers interviewed questioned the extent to which this was possible.

I mentioned feedback before written feedback. That's something that needs to be very consistent but that is very time consuming and I've tried it. I sat down in a handwriting lesson in a junior class and ERO had been mouthing off to us you have got to write down what the next steps are for the children and I sat down and wrote them written feedback; I had a queue out the door. (Brian)

I think one of the things we have been working on about formative assessment is that thing about written feedback in books. Having somewhere the children can see whatever is and it is not necessarily written down but it is supposed to be written down more than we do. It is more meaningful written down. There is a lot of data and a lot of information which says it should be written down. I think it is quicker for somebody to actually say it to you. It is all that positive written feedback which is terrific. It is one of my goals for this year is to write more. According to the ERO, the whole process doesn't take very long and we were talking about marking books and things like that and it is great the whole thing only takes you a minute but it doesn't – it doesn't take you just a minute. You might have eight children in that group, or eight children you are going to see that day just for that writing session, you are only going to deal with them. It is not always that easy. That is the reality. (Susan)

7.9. 2 ARBs, The Exemplars, NEMP tasks

Since the introduction to schools of the ARBs in 1996, there have been a number of assessment tools that have been developed to assess the curriculum. A number of these are used for assessment in literacy and numeracy such as STAR, asTTle and the Numeracy Project. These tools have been implemented in schools in a rapid and pervasive manner. For the three case study schools, the ARBs and The Exemplars (apart from Writing in years 1 to 3, which is not covered by asTTle) have largely been ignored or used very spasmodically. Teachers report that their focus on literacy and numeracy does not support the use of these tools. While this research focuses on the manageability and utility of assessment, not on questions of economy, it is perhaps timely to ask whether the money devoted to the development of these resources and the training of staff in their use achieved sufficient uptake to justify the outlay of resources. The same question could be asked about the proliferation in the development of tools for assessing mathematics with eight different ways to assess the subject being available in years 4 to 8. The schools largely reported use of NEMP tasks in a similar way to ARBs and the Exemplars; they were aware of them and believed they were good tools but either did not use them or used them only spasmodically in individual classrooms.

7.9.3 School Entry Assessment

SEA is another tool where use could be described as mixed. No school in the research was using the kit as directed; principals have little idea of how it is being operated and made very little use of the data. This was similar to other studies on the use of SEA (Dewar & Telford, 2003; Dixon & Williams, 2000; Goldring, 1999). SEA is one tool where it is possible to trace a direct relationship between problems related to manageability and the limited use made of the tool in schools.

The fate of SEA does illustrate some unique features of assessment in New Zealand schools. The MoE supplied a tool to all New Zealand schools and requested that the results be sent to them. The majority of schools, all of which by then had eight years of self-management, did not implement the tool in the way it was designed and ignored the request to send in the results. This lack of implementation and compliance was not officially followed up with individual schools. It would be surprising if other education systems operated in the same manner. One interpretation is that the non-reporting of SEA may illustrate the relative power given to individual schools to be self-managing under Tomorrow's Schools; if a central requirement does not fit with the choices made by a school then the requirement is ignored. An alternative explanation may have been that the Ministry concluded that it had tried to get schools to use a tool that they later realised was unmanageable and therefore ignored its lack of use and the non-reporting of the results.

7.9.4 asTTle

asTTle Writing has been used by Endeavour and Kupe for writing assessment since 2004, while Tasman plans to start its use this year. Schools that are using asTTle Writing also use the Exemplars for assessing writing at years 1 to 3. For Endeavour and Kupe, asTTle Writing was reported by all teachers as the assessment activity that was the most successful. They did acknowledge it took a lot of work to ensure there was consistent assessment when using this tool. Furthermore all respondents acknowledged this was a very time consuming form of assessment but felt the utility of the data outweighed concern about the time required for assessment and moderation. The only frustration that teachers reported with asTTle Writing was that the tool could not be used below year 4. This meant that teachers in year 3-4 composite classes needed to use two assessment tools in the same class.

Tasman is using asTTle reading and report finding the data very valuable. Endeavour is making more informal use of this assessment. Only Tasman was using asTTle Mathematics.

7.9.5 STAR

All schools were administering STAR twice a year and report positively on its utility and manageability. As stated earlier in terms of manageability, Tasman's administration of both STAR and asTTle Reading twice a year could be seen as challenging in terms of manageability. This means that the students are being subjected to four formal reading assessments a year, or the equivalent of one for every ten weeks of school.

7.9.6 Numeracy Project Assessment

Kupe is in the second year of the Numeracy Project and its assessment procedures are largely dictated by the Project. The staff at Endeavour and Tasman have encountered some problems with the Numeracy Project assessment in terms of the time required to undertake the assessment, clarifying their own expectations for how often and what form ongoing assessment should take, and inducting new staff who have not received the training in this form of teaching and assessment. Brian, the principal at Tasman is enthusiastic about the data the GLOSS assessment is providing to inform his teachers while Richard at Endeavour originally questioned the utility of the data for school-wide use.

Using the full diagnostic test of the Numeracy Project for assessment is similar to taking Running Records in that they are both individually administered assessments and therefore take a considerable amount of time. The quote from Ross, the AP at Tasman, illustrates both the challenge of the individual assessment requirements of the Numeracy Project and Running Records.

We doing the GLOSS twice a year and the sort of tick sheet things each different stage early additive and what they should be able to do, tick what they can do it. I could never get my head around it. I missed all the training so when I came back I did a little short course. So this is the first year I have got in up and running it is the same as the Running Records it's the manageability.

Where assessments take a large amount of time the consistency of application can be put at risk. As Susan from Endeavour stated:

The Numeracy Project assessment I suspect, in fact I know, is not being used in the school as it should be as it is set down in our guidelines, in fact I know it is not. I set myself a task I was going to retest all my child using the whole number assessment to see how long it took me before I launched into all the staff to say they should do it – to know how much time it took. Because I know what is going to happen because there will be a riot. The teachers will say we haven't got enough time to do that. I embarked on that last week and I have managed to do two of the whole assessment.

The Numeracy Project has an unusual approach to assessment. While a school is undertaking the project there is a very prescriptive approach to assessment. Schools are directed how and when to carry out the assessment. However, when the training is completed the trainers and advisers are very reluctant to give guidance on an appropriate schedule for on-going assessment, leaving such decisions for individual schools to work out for themselves. Both Tasman and Endeavour have changed their approach over the previous 18 months and the impression given is that both schools may change the schedule yet again.

7.9.7 Planning and Reporting Requirements

Interviews with teachers and principals demonstrated that the first two years of this requirement added little of value to schools and was largely an add-on which mainly concerned the principal and was irrelevant to the school's operations. However, this process developed in subsequent years and the setting of targets is now more likely to be done collectively with more ownership by the staff as a whole, and their achievement becoming more central to the school's operations. Recently published research by NZCER (Hipkins, Joyce & Wylie, 2007) on the implementation of planning and reporting has shown that the process is having a positive effect in primary schools.

7.9.8 Summary of Recent Developments – Research Question Three

One consideration that could be argued is that the proliferation in the development of assessment tools available to New Zealand schools in the last four years has presented problems almost as difficult as the lack of assessment tools fifteen years ago. In the early 1990s, schools had no tools to assess learning against the new curriculum but by

2003 there may have been too many. The results of this study do not dismiss this argument but the fact that all three schools give major priority to numeracy and literacy assessment has meant they have been able to make some clear choices about which tools to use without becoming taxed with manageability problems. However, as stated, there is a concern that some of these tools may be being administered too often.

The large amount of time Endeavour and Kupe give to moderating the assessment of writing does raise some questions of manageability and utility. While it is defensible to spend time getting agreement among teachers about the appropriate next steps for both individuals and cohorts of children, spending a great deal of time getting agreement on whether a sample of year 4 (eight year olds) writing is *level two basic*, or *level two proficient* is questionable for what is a very low stakes assessment.

Finally, a significant challenge to the manageability and utility at the school system level (the outer circle of figure 7.1) may be the reluctance of schools to abandon some of the less useful procedures developed in the past for assessing and recording student progress. This includes retaining the use of Running Records for competent readers, having multiple procedures for collecting and recording achievement information in terms of profiles, progress cards, and student folders, and gathering data in the non-core curriculum that is of questionable value.

CHAPTER EIGHT: LIMITATIONS OF THIS RESEARCH, RECOMMENDATIONS AND FURTHER RESEARCH

8.1 Limitations of this Research

Like all case study research, it is questionable practice to generalise from the context of three schools to other schools. This highly contextualised study has examined some major issues and drawn a number of conclusions. It remains for readers to decide to what extent the conclusions from this research can be applied to their own contexts.

When writing on the limitations of case study research, Hodkinson and Hodkinson (2001) state that the approach is strongest when the researcher's expertise and intuition are maximised but this raises doubts about objectivity. The fact that I have been a principal during the period under investigation provided a number of advantages in terms of the selection of areas of school practice and assessment to study. As stated at the beginning of this thesis, I have made a strong attempt to suspend my own position on aspects of assessment in this study. With these limitations in mind it is up to the reader to decide how objective the research has been.

Assessment in New Zealand schools is in a period of rapid change. It has been very challenging to evaluate what is a complex and moving body. From when this study began in mid 2004 to the time of writing (mid-2007 to April 2009), assessment issues have changed significantly. For this reason, it has been difficult to define the period under examination. The interviews were carried out over a period from November 2006 to May 2007 which in some ways defines the closing date of the research. This research therefore is a snapshot of the views, reflections and reported practice of 11 school staff (principals, senior management and teachers) in three Wellington schools during a six month period in 2006-2007. The postscript interviews of the principals did allow comment and description of changes in assessment practice in the individual schools two years later. However, these later interviews focused on issues that came out of the original school visits rather than matters relating to subsequent events such as the revised New Zealand Curriculum (MoE, 2007) and the approaching national standards. As stated above it is the role of the reader to decide to what extent this research, which is highly contextualised in time and place, applies to their own contexts.

As noted earlier, the original design of this research did call for more scale A teachers to be interviewed. Problems with staff turnover at Tasman and the decision to interview the assessment leaders at Kupe rather than scale A teachers meant that the research did not provide a strong perspective that reflected the views of basic scale teachers. The fact that all the focus group members were also principals did mean that the perspective of principals was the dominant voice in the research process. A design that included greater participation by classroom teachers (Scale A) may have provided variations on some research findings.

In a similar manner the interview process was the predominant method of data gathering in the research. This meant that the conclusions reached and reports given relied on the accuracy of the participants' recall.

In making such decisions on the credibility of the study and its applications to other contexts, readers rely on information that establishes the trustworthiness and transferability of the research. The investigation attempted to provide triangulation of data within the limitations of qualitative research design. These attempts included strategies to obtain corroboration for various data, for example, through the use of more than one case study school and interviewing between three and four participants in each school. The use of the postscript interviews and the focus groups were also aimed at providing corroborative triangulation. The use of a wide range of information sources such as documentation examination, semi-structured interviews, ERO reports and the literature review sought to build a comprehensive picture of assessment issues and provide a sense of coherence to the research (Hall, 2007).

The research supervisor critiqued the interview schedule, while the Human Ethics Committee of the University's College of Education approved the schedules in terms of their ethical implications. The transcripts of the taped interviews were returned to participants for checking to ensure that they were correct.

As stated above the need to set a closing date on the content of the study meant that some important current issues have not been covered, or covered very briefly. For example, the requirement that all schools purchase an electronic school management system with an assessment recording package is a major issue for schools and has serious implications for assessment and reporting practices in schools. This issue has been dealt with only briefly. While many of the interviews mentioned the key

competencies and the associated changes with the New Zealand Curriculum (MoE, 2007), these topics have not been addressed. In saying this, it needs to be acknowledged that the revised Curriculum is a far less complex document than the previous curriculum statements that have been critiqued in this thesis. However, it also needs to be noted that the changes have retained the levels and the general structure of the original curriculum. In discussing New Zealand assessment tools, both the recently revised PAT Mathematics and Reading tests have been discussed briefly but the development of E-asTTle has not been covered.

8.2 Recommendations

8.2.1 Non-Core Curriculum

Based on the data from the case study schools, it could be argued that much of the assessment and reporting that is undertaken in the areas of health and physical education, social studies, science, technology, the arts, oral language and Te Reo is limited in its nature and contributes very little to subsequent teaching and valid reporting. It would be appropriate for stakeholders such as the Ministry of Education, ERO, NEMP, NZEI, NZPF and STA to work towards gaining a consensus in relation to what is appropriate assessment for schools to undertake in these areas. The fact that it is extremely difficult to assess in all of these areas to an appropriate depth is evident in the work of both NEMP and ERO. NEMP only assesses three areas of the curriculum each year, while ERO reviews two or three areas of government interest every six months. The rationale for NEMP in only assessing a curriculum area every four years is that it requires such a time period to identify significant trends in achievement. This same rationale can be applied to individual schools in their tracking of students' learning for school reporting. In terms of areas of government interest and the associated work of ERO, it is possible that the limiting of analysis to two or three areas each year is for economic reasons - it is probably unnecessary and inefficient for this large organisation to cover all areas on an annual basis. The same rationale, if it is correct, does not apply to schools. The NAGs and ERO reports contain clear expectations that schools should assess and record in all those areas annually. A change to allow schools to be more selective about which areas they assess in a meaningful way on a similar rotation system may mean that more use is made of The Exemplars and other tools. If the three schools in this investigation in any way reflect

wider national use of the Exemplars, a question exists over the uptake of this tool given the outlay involved in their development.

8.2.2 The Written Recording of Formative Assessment Feedback

Comments

While it is hard to argue against the adoption of the feedback elements of formative assessment, how this feedback is given may require further examination. Both ERO and school leaders may need to question whether expecting written comments for most of these interactions is manageable and whether it may be counterproductive by taking away time from teaching. Two of the teachers in the study identified the recommendation of ERO for written formative comments in students' work as being impractical when teachers attempt this practice in all areas of teaching.

8.2.3 The Frequency of Assessment

A further action that could assist schools in the manageability of assessment in New Zealand primary schools is for test developers, the Ministry of Education, professional development advisers, and ERO to provide schools with more guidance on how frequently assessment should take place. Timperley (2003) criticised how often Running Records were being administered in the schools that she studied. She also made the point that these assessments were inappropriate for the level of children. In terms of frequency, more guidance would address manageability issues regarding the use of such tools as STAR, asTTle and the Numeracy Project. The Principal of Endeavour reported that he knew of a school that administered STAR three times a year. As noted previously, Tasman's administration of both STAR and asTTle Reading twice a year seems unnecessary. This means that the students are being subjected to four formal reading assessments a year, or the equivalent of one for every ten weeks of school. It should be noted that the tests themselves, either through their manuals or technical reports, give no guidance on how often they should be administered. Guidance is also absent from the MOE publications such as *What Matters Most* and *Curriculum Update*. This is an issue in terms of manageability.

There is also an issue with the complex assessment associated with the Numeracy Project. Both Endeavour and Tasman Schools changed their assessment routine for the Numeracy Project twice in the last three years and gave the clear impression that another change was being considered. More guidance on the nature and frequency of assessment would be very helpful to schools.

8.2.4 The Economics of Assessment Development in New Zealand Schools

While this research investigated the manageability and utility of assessment rather than the economics of the development of assessment tools, some questions do need to be raised in this area. From having no, or very few, assessment tools in the early nineties New Zealand would now seem to have more instruments than needed. As noted in Chapter Four, there are now eight ways of assessing children's mathematics in New Zealand. The tools developed in the last ten years include: ARB, The Exemplars, STAR, asTTle, and the revised PAT Mathematics and Reading. These have all been developed directly through government contracts (asTTle, ARB, Exemplars) or through BOT funds being used to purchase STAR and PAT. It could be argued that there is now considerable duplication in the system. STAR, PAT Reading and asTTle Reading are used for very similar purposes, as are asTTle and PAT Mathematics; and an argument can be made that the Exemplars and the ARBs perform similar functions in our schools. It would be surprising if other education systems serving similar populations to New Zealand have spent taxpayers' money on developing multiple assessment tools while leaving decisions on their use to individual schools. This issue has implications for the prudent use of educational funds and resources, including those devoted to professional development for training school staff in how to administer and use the tools.

8.3 Suggestions for Further Research

8.3.1 On-going Research on Assessment Issues in New Zealand Schools

As the assessment landscape keeps changing there is a need for on-going research on the implementation of assessment in New Zealand schools. While all areas of assessment would benefit from investigation there are a number of elements of the

curriculum where particular scrutiny is appropriate. This includes the Numeracy Project which is a major change in teaching practice. While the Project has been subject to monitoring on a nationwide basis (Annan, 2006) an examination of how it is being implemented, especially in the area of assessment at the individual school and teacher level, would be illuminating.

The present research highlighted issues in assessing the non-core curriculum, for example the arts, physical education, health, Te Reo, social studies, science and technology. Further research is needed in the assessment and teaching of these areas to identify what represents a manageable and useful regime.

On a more general note, research on assessment practices in New Zealand schools needs to provide greater information, both quantitative and qualitative. Quantitative research will enable information on widespread practices to be gathered. Qualitative research will provide greater depth of understanding of assessment practices both as a trigger for quantitative surveys and as follow-up when quantitative surveys suggest something needs deeper exploration.

A final word on manageability is pertinent in order to conclude this research. Prior to the introduction of the education reforms, new developments in curriculum and assessment were carefully trialled before being introduced for widespread use in schools. It is now clear that the reforms associated with Tomorrow's Schools, The New Zealand Curriculum Framework, the work of ERO, and the reporting of assessment by schools, were not worked through with anything like the attention to the possible implementation problems that schools later encountered. The conceptualisation in this thesis of "three waves" of assessment activity provides a description of the major adjustments that were needed systematically in New Zealand schools to cope with the reform process.

REFERENCES

- Abbot, D., Broadfoot, P., Croll, T., & Pollard, A. (1994). Some sink, some float: National Curriculum Assessment and accountability. *British Educational Research Journal*, 20(2), 157-174.
- Absolum, M. (2006). *Clarity in the classroom: Using formative assessment-building classroom relationships*. Auckland, New Zealand: Hodder Education.
- Aikin, S. (1994). Primary problems and the New Zealand Curriculum Framework. *New Zealand Annual Review of Education*, 4, 57-76.
- Allen, P., Crooks, T., Hearn, S., & Irwin, K. (1997). *Te Tiro Hou: Report of the Qualifications Framework Inquiry*. Wellington, New Zealand: PPTA.
- Alphonse, N. (1999). Teachers, educational reforms and the credo of managerialism: Form professionals to pedagogic technicians. *Delta*, 15(2), 13-26.
- Annan, B., Ell, F., Fisher, J., Hawera, N., Higgins, J., Irwin, K., Tagg, A., Thomas, G., Trinick, T., Ward, J., & Young-Loveridge, J. (2007). *Findings from the New Zealand Numeracy Development Projects 2006*. Wellington, New Zealand: Learning Media.
- Auckland Uniservices Ltd. (1999). *The use of exemplars in outcomes-based curricula: An international review of the literature. Final Report to the Ministry of Education*. Auckland, New Zealand: University of Auckland.
- Baker, E. L., Linn, R.L., & Herman, J.L. (1996). *CRESST: A continuing mission to improve education assessment*. Los Angeles: UCLA Centre for the Study of Evaluation and the National Centre for Research on Evaluation, Standards and Student Testing.
- Baker, F. (1997). *Developing portfolios for assessment. A practical guide for teachers*. Auckland, New Zealand: Kohia Teachers Centre.
- Baker, F., & Lorrigan, G. (1995). *Assessment: Putting the big picture together*. Auckland, New Zealand: Kohia Teachers Centre.
- Baker, F., & Lorrigan, G. (1996). *Assessment ways forward*. Auckland, New Zealand: Kohia Teachers Centre.
- Beck, L. R., & St.Gerorge, R. (1983). The alleged cultural bias of PAT: Reading Comprehension and Reading Vocabulary Tests. *New Zealand Journal of Educational Studies*, 18(1), 32-47.

- Black, P., & Wiliam, D. (1998). Assessment and classroom learning: Theory and practice. *Assessment in Education: Principles, Policy & Practice*, 5(1), 131-138.
- Black, P., & Wiliam, D. (1998). *Inside the black box*. London: NFER Nelson.
- Black, P., & Wiliam, D. (2006). Developing a theory of formative assessment. In Gardner, J. (Ed.). *Assessment and Learning*. London: Sage. 47-61.
- Blaiklock, K. E. (2004). A critique of Running Records of children's oral reading. *New Zealand Journal of Educational Studies*, 39(2), 241-253.
- Boyd, S. (2003). Assessment Resource Bank in the assessment mix. *SET Research information for teachers*, 2003(2), 28-34.
- Boud, D. (1986). *Implementing student self-assessment*. Green Guide No. 5. Sydney, Australia: Higher Education Research and Development Society of Australia.
- Bridges, S. J. (1992). *Working in Tomorrow's Schools: Effects on primary teachers - A Christchurch study*. (Report No 92-3). Christchurch, New Zealand: University of Canterbury, Education Department.
- Broadfoot, P. (1995). Performance assessment in perspective: International trends and current English experience. H. Torrance (Ed.), *Evaluating authentic assessment: Problems and possibilities in new approaches to assessment*. (pp 9-43). Buckingham, U.K: Open University Press.
- Broadfoot, P. (2004). Redefining assessment. The first ten years of education assessment. *Education Assessment*, 11(1), 7-26.
- Broadfoot, P. (2007). *An introduction to assessment*. New York: Continuum International.
- Brody, L. E., & Lane, S. (1995). Progressive Achievement Test of Mathematics [Revised] [Electronic version]. *Mental Measurement Yearbook*, 12.
- Brown, G. (2000). Information skills: How well can New Zealand students find information? *SET Research information for teachers*, 2000(1) 26-31.
- Brown, G. T. L. (2002). *Teacher's conceptions of assessment*. Unpublished doctoral thesis, University of Auckland, Auckland, New Zealand.
- Burns, R. B. (2000) *Introduction to research methods* (4th ed.). Australia: Pearson Education.

- Butterworth, G. (1998). *Reforming education: The New Zealand experience*. Ministry of Education. Palmerston North, New Zealand: Dunmore Press.
- Capper, P. (1991). Curriculum 1991. *New Zealand Annual Review of Education*, 1, 15-27.
- Chamberlain, M. (2000). *The national assessment strategy and the relationship between teaching, learning and student achievement*. Paper presented at the National Assessment Regional Seminars, Wellington, New Zealand.
- Chamberlain, M. (2001). *The development of the exemplars in New Zealand: Background and rationale* [Electronic version]. *TKI Assessment readings and research*. Retrieved July 8, 2007, from http://www.tki.org.nz/r/assessment/research/mainpage/research_ref_e.php
- Clark, S. (2005). *Formative assessment in action*. London: Hodder Murray.
- Clark, S., Timperley, H., & Hattie, J. (2003). *Unlocking formative assessment*. Auckland, New Zealand: Hodder Moa Beckett.
- Clay, M. M. (1985). *The early detection of reading difficulties*. Auckland, New Zealand: Heinemann.
- Clay, M. M. (1993). *Reading recovery*. Auckland, New Zealand: Heinemann.
- Clay, M. M. (2002). *An observation survey of early literacy achievement* (2nd ed.). Auckland, New Zealand: Heinemann.
- Codd, J. (1999). Education reform, accountability and the culture of distrust. *New Zealand Journal of Education Studies, Special Issue: A decade of reform in New Zealand education: Where to now*, 45-53.
- Codd, J., Harker, R., & Nash, R. (Eds.). (1990). *Political issues in New Zealand Education*. Palmerston North, New Zealand: Dunmore Press.
- Codd, J., McAlpine, D., & Poskitt, J. (1995). Assessment policies in New Zealand: Education reform or political agenda. In R. T. Peddie, & B. Tuck (Eds.), (pp32-63). *Setting the standards: Issues in assessment for national qualifications* Palmerston North, New Zealand: The Dunmore Press.
- Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education*. (5th ed.). London: Routledge Farmer.
- Connor, C. (1991). *Assessment and teaching in the primary school*. London: Falmer Press.
- Consumer. (June 2002). Measuring your kids' progress: PATs in perspective. *Consumer*, 415, 18-21.
- Cowie, B., & Bell, B. (1999). A model of formative science education. *Assessment in Education*, 6(1), 101-117.

- Creswell, J. W. (2005). *Education research planning, conducting, and evaluating quantitative and qualitative research* (2nd ed.). Upper Saddle River, NJ: Pearson/Merrill/ Prentice Hall.
- Creswell, J. W., & Plano-Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.
- Critchely-Charlton, B. (2005). *Informal assessment strategies*. Ontario, Canada: Pembroke Publishers.
- Croft, A. C., Strafford, E., & Mappa, L. (2000). *Stocktake/evaluation of existing diagnostic tools in literacy and numeracy in English*. Wellington, New Zealand: NZCER.
- Croft, A. C., Dunn, K., & Brown, G. (2000). *Essential skills assessment: Teachers Manual*. Wellington, New Zealand: NZCER.
- Croft, C. (1984). PAT: Study Skills - after the tests, what next? *SET Research information for teachers*, 1984(2), (item 4)
- Croft, C. (1999). *School-wide assessment: Using the assessment resource banks*. Wellington, New Zealand: NZCER.
- Croft, C. (2001). *Resource Bank in English*. Paper presented at the 27th Annual Conference for the International Association for Education Assessment (IEA), Rio de Janeiro, May 6-11, 2001.
- Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of Education Research*, 58, 438-481.
- Crooks, T. (2002). Education assessment in New Zealand schools. *Assessment in Education*, 29, 237-254.
- Crooks, T., & Flockton, L. (2004). *National education monitoring project reading and speaking assessment results*. Dunedin, New Zealand: University of Otago, Education Assessment Research Unit,
- Daniels, J. C., & Diack, H. (1958). *Standard reading test*. London: Hart-Davis Educational.
- Darr, C. (2003). asTTle: An early exploration of new technology for teachers. *SET Research information for teachers*, 2003(1), 10-14.
- Darr, C., Neill, A., & Stephanou, A. (2006). *Teachers Manual: Progressive Achievement Test: Mathematics*. Wellington, New Zealand: NZCER.
- Darr, C., McDowall, S., Ferral, H., Twist, J., & Watson, V. (2008). *Teachers Manual: Progressive Achievement Test: Reading*. Wellington, New Zealand: NZCER.

- Dearing, R. (1994). *The national curriculum and its assessment: Final Report*. London: School of Curriculum and Assessment Authority.
- Denzin, N. K. (1998). Emancipatory discourses and the ethics and politics of interpretation. In N.K. Denzin, & Y. Lincoln (Eds.), *Collecting and interpreting qualitative materials* (pp. 435—472). Thousand Oaks, Ca: Sage.
- Dewar, S., & Telford, M. (2003). *A study of the use of school entry assessment (SEA) in schools*. Wellington, New Zealand: Ministry of Education.
- Dixon, D. (1999). *The effects of policy on practice: An analysis of teachers' perceptions of school-based assessment practice*. Unpublished M.Ed Thesis, Massey University, Palmerston North, New Zealand.
- Dixon, H., & Williams, R. (2000). School entry assessment: The implementation issues. *SET Research information for teachers, 2000* (1), 32-36.
- Doig, B. (2006). Large-scale mathematics assessment: Looking globally to act locally. *Assessment in Education, 1* (3), 265-288.
- Donnelly, K. (2006). *The New Zealand Curriculum: A submission on the draft for consultation 2006*. Wellington, New Zealand: Education Forum.
- Duffy, B. (1999). The analysis of documentary evidence. In B. Bell, (Ed.), *Doing your research project* (pp. 103-126).. Buckingham, UK: Open University Press.
- Dunn, K. (2001). Essential skills assessments: How well can students read between the lines. *SET Research information for teachers, 2001* (2), 24-26.
- Dunn, K., Trafford, E., & Marston, C. (2003). *Classroom assessment practices in English and mathematics at years 5, 7, and 9*. Wellington, New Zealand: NZCER.
- Education Review Office. (1994a). *Good New Zealand schools*. Wellington, New Zealand: Education Review Office.
- Education Review Office. (1994b). *Mathematics in the New Zealand Curriculum*. Wellington, New Zealand: Education Review Office.
- Education Review Office. (1995a). *Arts in the New Zealand Curriculum*. Wellington, New Zealand: Education Review Office.
- Education Review Office. (1995b). *Assessing student achievement*. Wellington, New Zealand: Education Review Office.
- Education Review Office. (1996). *Reporting student achievement*. Wellington, New Zealand: Education Review Office.

- Education Review Office. (1998). *Assessing children's curriculum achievement*. Wellington, New Zealand: Education Review Office.
- Education Review Office. (1999). *Student assessment: Practices in primary schools*. Wellington, New Zealand: Education Review Office.
- Education Review Office. (2000). *In time for the future: A comparative study of mathematics and science education*. Wellington, New Zealand: Education Review Office.
- Education Review Office. (2001). *The New Zealand Curriculum: An ERO Perspective*. Wellington, New Zealand: Education Review Office.
- Education Review Office. (2004). *St Joseph's School Upper Hutt. ERO Review*. Wellington, New Zealand: Education Review Office.
- Education Review Office. (2007). *Assessing and feeding back (PROCESS INDICATORS)*. Retrieved August 17, 2007, from <http://www.ero.govt.nz/EdRevInfo/Schedrevs/AssessFeedPI.doc>
- Eisner, E. W. (1999). The use and limits of performance assessment. *Phi Delta Kappan*, 80(9), 658-661.
- Ell, F. (2001). *Mathematics in the New Zealand Curriculum - A concept map of the current document*. Auckland, New Zealand: University of Auckland, Project asTTle.
- Elley, W. B. (1985). *Lessons learnt about LARIC*. Christchurch, New Zealand: University of Canterbury, Department of Education.
- Elley, W. B. (1993). Curriculum reform: Forwards or backwards. *New Zealand Annual Review of Education*, 3, 37-50.
- Elley, W. B. (1996). Curriculum reform: Forwards or backwards. *Delta*, 48 (1), 11-18.
- Elley, W. B. (1997). *An evaluation of Alan Duff's "Books in homes" programme. Final Report*. Wellington, New Zealand: Ministry of Education.
- Elley, W. B. (2001). *STAR: Supplementary tests of achievement in reading*. Wellington, New Zealand: NZCER.
- Elley, W. B. (2004). Curriculum reform in retrospect. In A.M. O'Neil, J. Clark, & R. Openshaw (Eds.), *Reshaping culture, knowledge and learning. Policy and content in the New Zealand Curriculum Framework* (Vol. 1). Palmerston North, New Zealand: Dunmore Press.
- Elley, W. B. (2005). *Facts and fallacies about standards-based assessment [Electronic Version]*. Retrieved May 18, 2007 from www.mcleans.school.nz/news/page/2005/ncea/htm.

- Elley, W. B. (2006). How well do our students achieve? A summary of 35 years of international comparisons. *SET Research information for teachers*, 2006(1), 2-8.
- Elley, W. B., & Reid, N.A. (1969). *Progressive achievement tests of reading: Teacher's manual*. Wellington, New Zealand: NZCER.
- Elley, W. B., Hall, C., & Marsh, R. (2004). Rescuing NCEA: Some possible ways forward. *New Zealand Annual Review of Education*, 14, 5-26.
- Ewing, J. L. (1970). *The development of the New Zealand primary school curriculum 1877-1970*. Wellington, New Zealand: NZCER.
- Fancy, H. (2004). *Education reform: The New Zealand experience*. Paper presented at the Education for change symposium, Melbourne, Australia.
- Fancy, H. (2004). Presentation to ERO staff, 9 June, 2004, Brentwood Hotel Wellington, New Zealand. [Electronic Version].
- Fancy, H. (2006). *Making a good education system great*. Paper presented at The Correspondence School Parents and Supervisors Association Conference, Portland Towers, Wellington, New Zealand. 18 May 2006.
- Fawson, P. C., Ludlaw, B. C., Reutzel, D. R., Sudweek, R., & Smith, J. A. (2006). Examining the reliability of Running Records: Attaining generalised results. *The Journal of Education Research*, 100 (2), 113-128.
- Flockton, L. (1999). *School-wide assessment: National Education Monitoring Project*. Wellington, New Zealand: Learning Media.
- Flockton, L., & Crooks, T. (1999). *Assessment results 1999: Science. Art. Graphs, Tables and Maps*. Dunedin, New Zealand: University of Otago, The Educational Assessment Research Unit.
- French, A. (2000). *The heart of the matter: How the Education Review Office evaluates pre-tertiary education*. Wellington, New Zealand: Victoria University, Victoria Link.
- Fullan, M. (2005). *Leadership and sustainability*. Thousand Oaks, Ca: Corwin Press.
- Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Education research: An introduction*. Boston: Pearson Education.
- Gardner, J. (Ed.). (2006). *Assessment and learning*. London: Sage.
- Gerritsen, J. (2000). New assessment initiatives. *Education Gazette*, 79 (14)
- Gilbert, J. (2005). *Catching the knowledge wave? The knowledge society and the future of education*. Wellington, New Zealand: NZCER.

- Gilmore, A., Croft, C., & Reid, N. (1981). Burt Word Reading Test. Wellington, New Zealand: NZCER.
- Gilmore, A. (2002). Large scale assessment and teachers education capacity. Learning opportunities for teachers in the National Education Monitoring Project in New Zealand. *Assessment in Education*, 9 (3), 343-318.
- Gilmore, A., & Hattie, J. (2000). *Evaluation of the assessment resource banks - Final Report*. University of Canterbury, University of Auckland.
- Gilmore, A., & Hattie, J. (2001). Understanding usage of an internet based information resource for teachers: The Assessment Resource Bank. *New Zealand Journal of Educational Studies*, 36 (2), 237-257.
- Gipps, C. V. (1994). *Beyond testing: Toward a theory of educational assessment*. London: Routledge Farmer.
- Gipps, C. V. (1995). Reliability, validity and manageability in large-scale performance assessment. In H. Torrance (Ed.), *Evaluating authentic assessment* (pp, 105-123). Buckingham, UK: Open University Press.
- Glaser, B. G. (2000). The future of grounded theory. *Grounded Theory Review*, 1, 1-8.
- Glasswell, K., & Brown, G. T. L. (2003). *Accuracy in scoring of writing: Study in large scale scoring of asTTle Writing assessment*. Auckland, New Zealand: University of Auckland/Ministry of Education.
- Glasswell, K., Parr, J., & Aikman, H. (2001). *Developing the asTTle rubrics for scoring extended writing tasks*. Auckland, New Zealand: University of Auckland/Ministry of Education.
- Goldring, M. (1999). *School entry assessment practices in schools*. Christchurch, New Zealand: University of Canterbury, University of Canterbury.
- Goodman, Y. M., & Burke, C. L. (1972). *Reading miscue inventory: Manual and procedures*. New York: MacMillan.
- Gordon, L. (1991). The bulk funding of teachers salaries. *New Zealand Annual Review of Education*, 1, 28-58.
- Gordon, L. (1996). School choice and the quasi-market in New Zealand: 'Tomorrow's Schools' today. *Oxford Studies in Comparative Education*, 6 (1).
- Gordon, L., & Whitney, G. (1997). Giving the 'hidden hand' a helping hand? The rhetoric and reality of neoliberal education reform in England and New Zealand. *Comparative Education*, 33, 453-467.
- Graves, D. H. (1983). *Writing: Teachers and children at work*. Exeter, NH: Heinemann.

- Greaney, K. (2004). Factors affecting PAT Reading Comprehension performance: A retrospective analysis of some year 4-6 data. *New Zealand Journal of Education Studies* 39 (1), 3-21.
- Hall, C. (2005). The National Certificate of Education (NCEA): Is there a third way? In J. S. Codd. & K. Sullivan (Eds.), *Education Policy Directions in Aotearoa New Zealand*. Wellington, New Zealand: Dunmore Press.
- Hall, C. (2006). *Planning the assessment for programmes and courses: A guide for tertiary level educators*. [Unpublished lecture handout]. Victoria University of Wellington, New Zealand: School of Educational Psychology and Pedagogy.
- Hall, C. (2007). *Outcomes-based education*, [Unpublished lecture handout]. Victoria University of Wellington, School of Education Studies.
- Hall, C. (2007b) *Evaluation as a method of educational enquiry*, [Unpublished lecture handout]. Victoria University of Wellington, School of Education.
- Hall, K., Webber, B., Varley, S., Young, V., & Dorman, P. (1997). A study of teaching and assessment at key stage one. *Cambridge Journal of Education* 27(1), 107-123.
- Ham, V., Findlay, G., Schwier, R., & Davies, G. (2000). *Evaluation of the Assessment Resource Banks as an electronic resource*. Christchurch, New Zealand: Christchurch College of Education.
- Harrison, M. (2004). *Education matters: Government, markets and New Zealand schools*. Wellington, New Zealand: Education Forum.
- Hattie, J., & Peddie, R. (2003). School reports: Praising with faint damns. *SET Research information for teachers*, 2003(3), 4-9.
- Hattie, J. A. C., & Brown, G. T. L. (2004). *Cognitive processes in asTTle: The SOLO taxonomy*. Auckland, New Zealand: University of Auckland/Ministry of Education.
- Hattie, J. A. C., Brown, G. T. L., Keegan, P., Irving, S. E., McKay, A. J., & Sutherland, T. (2003). *Validation evidence of asTTle Reading assessment results: Norms and criteria*. Auckland: University of Auckland/ Ministry of Education.
- Hattie, J. H. (1999). *Influences on student learning*, [Inaugural Lecture: Professor of Education] Auckland, New Zealand: University of Auckland.
- Heron, M., Hucker, J., Rooney, S., Robinson, V., Lai, M. K., & Morse, K. (2001). Progressive Achievement testing at Southern Cross Middle School. *SET Research information for teachers*, 2001 (2), 35-38.

- Higgins, J. (2002). The Numeracy Development Project: Policy to practice. *New Zealand Annual Review of Education*, 12, 157-176.
- Hill, M. (1995). Self assessment in primary schools: A response to student questions. *Waikato Journal of Education*. 1, 61-70.
- Hill, M. (1999). Assessment in self-managing schools: Primary teachers balancing learning and accountability demands in the 1990s. *New Zealand Journal of Educational Studies*, 34(2), 176-185.
- Hill, M. (2000). *Remapping the assessment landscape: primary teachers reconstructing the assessment in self managing schools*. Unpublished doctoral thesis, University of Waikato, Hamilton, New Zealand.
- Hill, M. (2000a). Dot, slash, cross: How assessment can drive teachers to ticking instead of teaching. *SET Research information for teachers*, 2000 (1), 21-25.
- Hill, M. (2003). Assessment in the real world - what teachers actually do. Wellington, New Zealand: *NZEI Occasional Paper*.
- Hipkins, R., Joyce, C., & Wylie, C. (2007). *School planning and reporting in action: The early years of the new framework*. Wellington, New Zealand: NZCER.
- Hodkinson, H., & Hodkinson, P. (2001).. *The strengths and limitations of case study research*. Paper presented at the learning and skills development agency conference, Cambridge, U.K on 5 Dec, 2001.
- Holton, D., Spicer, T., Thomas, G., & Young, S. (1996) *The benefits of problem solving in learning mathematics*. Final Report to the Ministry of Education. Wellington, New Zealand: Ministry of Education, Research and International Section.
- Hopkins, C. D., & Antes, R. L. (1990). *Classroom measurement and evaluation*. Itasca, IL: Peacock.
- Huberman, A. M., & Miles, M. B. (2002). *The qualitative researchers companion*. Thousand Oaks, Ca: Sage.
- Hughes, D. C. (1982). School Records: Some comments on recent proposals. *New Zealand Journal of Educational Studies*, 17 (2), 185-189.
- James, M. (2006). Assessment, teaching and theories of learning In Gardner, J. (Ed.). *Assessment and Learning*. London: Sage. 47-61.
- Johnson, R. L. (2004). Essential Skills Assessments: Information Skills [Electronic version]. *The Mental Measurement Yearbook*, 15. Retrieved, June, 24, 2007.
- Joyce, C., & Darr, C. (2004). Dusting off the teachers' manual. *SET Research information for teachers*, 2004 (3), 55-56.

- Kelly, F. (2000). *Ministry of Education policy interest in assessment*. Paper presented at the Examining Assessment Conference, New Zealand Council for Education Research.
- Khatttri, N., Reeve, A. L., & Kane, M. B. (1998). *Principles and practices of performance assessment*. Florence, Kentucky: Lawerence Erlbaum Associates
- Knight, N. (2000). Assessment for accountability, reporting and improving learning: Have we got the balance right? *Delta*, 2 (2), 39-62.
- Kruegar, R. (1998). *Focus groups: A practical guide for applied research*. Newbury Park, Ca: Sage.
- Lai, M. K., & McNaughton, S. (2003). *Profile of reading comprehension across a cluster of schools in Mangere, South Auckland*. Paper presented at the joint conference of Australian/New Zealand Association for Research in Education, Auckland, New Zealand.
- Lange, D. (1998). *Tomorrow's schools: The reform of education administration in New Zealand*. Wellington, New Zealand: Government Printer.
- Larson, V. (2005). NCEA: Can they fix it? *North and South*. (October, 2005), 42-55.
- Lauder, H., Hughes, D., Watson, S., Waslander, S., Thrupp, M., & Dupuis, A. (1999). *Trading in futures: Why markets in education don't work*. Buckingham, UK: Open University Press.
- Le Metais, J. (2002). *New Zealand curriculum stocktake: An international critique*. Wellington, New Zealand: Ministry of Education.
- Lee, J. J. (1980). Defining the curriculum. In G. H. Robinson, B. T. O'Rourke (Eds.), *Schools in New Zealand society*. (pp. 117-132) Auckland, New Zealand: Longman Paul.
- Limbrick, L., Keenan, J., & Girven, A. (2002). *Mapping the English Curriculum*. asTTle Technical Report 4. Auckland, New Zealand: Auckland College of Education.
- Lin, R. L., & Gronlund, N. E. (2000). *Measurement and assessment in teaching* (8th ed.). Upper Saddle River; NJ: Merrill.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, Ca: Sage.
- Lindvall, C. M., & Nitko, A. J. (1975). *Measuring student achievement and aptitude*. New York: Harcourt, Brace Javanovich.
- Livingstone, I. D. (1994). *The workloads of primary school teachers: A Wellington region survey*. Wellington, New Zealand: Chartwell Consultants.

- Lovett, S., & Sinclair, L. (2005). *The socialisation of teachers into a culture of assessment*. Wellington, New Zealand: NZCER.
- Macfie, R. (2008). Classroom controversy. *New Zealand Listener*. (February 9, 2008), 19-24.
- Madaus, G., & O'Dwyer, L. (1993). A short history of performance assessment. *Phi Delta Kappan* (80), 688-695.
- Mallard, T. (2000). *New Zealand Schools: A report on the compulsory schools sector in New Zealand 2000*. Wellington, New Zealand: Ministry of Education.
- Mallard, T. (2003). *A report on the compulsory schools sector in New Zealand. Minister of Education*. Wellington, New Zealand: Ministry of Education.
- Manno, B. V. (2004). Outcome-based education: Miracle cure or plague [Electronic Version]. *Hudson Institute Briefing Paper*, 165.
- Marshall, J., Coxton, E., Jenkins, K., & Jones, A. (2000). *Politics, policy, and pedagogy: Education in Aotearoa/New Zealand*. Palmerston North, New Zealand: Dunmore Press.
- Marston, C., & Kenneally, R. (2001). Student interest in science: Analysis of data from ARB. *SET Research information for teachers*, 2001 (3), 4-6.
- Maykut, P. & Morehouse, R. (1994). *Beginning qualitative research. A philosophical and practical guide*. London: Farmer Press.
- McCallum, B., Gipps, C., McAlister, S., & Brown, M. (1995). National Curriculum assessment: emerging models of teacher assessment in the classroom. In H. Torrance, (Ed), *Evaluating authentic assessment* (pp88-104) . Buckingham, UK: Open University Press.
- McMahon, M. (2000). *Evidence to action*. Paper presented at the National Assessment Regional Seminars. New Zealand.
- McMaster, A., & Bonallack, J. (1998). *Easy assessment in social studies*. Wellington, New Zealand: Jam Publications.
- Meagher-Lundberg, P., & Brown, G. T. L. (2001). *Item signature study: Report of the characteristics of reading texts and items from calibration 1*. Auckland, New Zealand: University of Auckland, Project asTTle.
- Meisels, S. J., DiPrima Bickel, D., Nicholson, J., Xue, Y., & Atkins-Burnett, S. (2001). Trusting teachers' judgments: A validity of a curriculum-embedded performance assessment in kindergarten to grade 3. *American Educational Research Journal*, 38 (1), 73-95.

- Mendelovits, J., Farkota, J., & Lindsay, J. (2000). *Evaluation of the New Zealand Assessment Resource Banks' methodology*. Melbourne, Australia: Australian Council for Education Research.
- Merriam, S. B. (1998). *Case study research in education: A qualitative approach*. San Francisco: Jossey-Bass Publishers.
- Mertens, S. B. (1997). *Research methods in education and psychology*. Thousand Oaks, Ca: Sage Publications.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessment. *Education Researcher*, 23, 13-23.
- Miller, M.D., Linn, R.L., & Gronlund, N.E. (2009) *Measurement and assessment in teaching*. Upper Saddle River, NJ: Pearson Education.
- Ministry of Education. (1992). *Mathematics in the New Zealand Curriculum*. Wellington, New Zealand: Learning Media.
- Ministry of Education. (1993a). *New Zealand Curriculum Framework*. Wellington, New Zealand: Learning Media.
- Ministry of Education. (1993b). *Science in the New Zealand Curriculum*. Wellington, New Zealand: Learning Media.
- Ministry of Education. (1993c). National Education Guidelines. *Education Gazette* (April 1993).
- Ministry of Education. (1994a). *English in the New Zealand Curriculum*. Wellington, New Zealand: Learning Media.
- Ministry of Education. (1994b). *Assessment: Policy to practice*. Wellington, New Zealand: Learning Media.
- Ministry of Education. (1995). *Technology in the New Zealand Curriculum*. Wellington, New Zealand: Learning Media.
- Ministry of Education. (1997a). *School Entry Assessment*. Wellington, New Zealand: Learning Media.
- Ministry of Education. (1997b). *Planning and assessment in English*. Wellington, New Zealand: Learning Media.
- Ministry of Education. (1997c). *Report of the mathematics and science task force*. Wellington, New Zealand: Ministry of Education.
- Ministry of Education. (1997d). *Governing and managing New Zealand schools: A guideline for boards of trustees. Part One: The National Education Guidelines*. Wellington, New Zealand: Learning Media.

- Ministry of Education. (1998). *Assessment for success in primary schools*. Wellington, New Zealand: Learning Media.
- Ministry of Education. (1999a). *Health and Physical Education in the New Zealand Curriculum*. Wellington, New Zealand: Learning Media.
- Ministry of Education. (1999b). National Education Administration Guidelines *Education Gazette*, 79, (29 November).
- Ministry of Education. (1999c). *Report of the Literacy Taskforce*. Wellington, New Zealand
- Ministry of Education. (1999d). *Information for better learning*. Wellington, New Zealand: Ministry of Education.
- Ministry of Education. (2000a). *The Arts in the New Zealand Curriculum*. Wellington, New Zealand: Learning Media.
- Ministry of Education. (2000b). *Sharpening the Focus*. (Issue 3, August 2000).
- Ministry of Education. (2000c). *Using Running Records: A resource for New Zealand classroom teachers*. Wellington, New Zealand: Learning Media.
- Ministry of Education. (2000d). *Report to the Ministry of Education: A review of the roles and responsibilities of the Education Review Office*. Wellington, New Zealand.
- Ministry of Education. (2001b). *Curriculum Update*, February, 2001(45).
- Ministry of Education. (2001d). *Education Standards Act - 2001*. Retrieved. June, 18, 2007 from <http://www.minedu.govt.nz/index.cfm?layout=document&documentid=6426&CFID=5858805&CFTOKEN=33605423>.
- Ministry of Education. (2002a). *Curriculum stocktake report*. Wellington, New Zealand.
- Ministry of Education. (2002b). *Sharpening the focus: A newsletter for boards of trustees, principals and teachers on the revised national guidelines* (8). Wellington, New Zealand.
- Ministry of Education. (2003a). *The New Zealand Curriculum Exemplars*. Wellington, New Zealand: Learning Media and The Learning Centre Trust of New Zealand.
- Ministry of Education. (2003b). Analysing student achievement data. *Planning for Better Student Outcomes*, September 2003.
- Ministry of Education. (2003c). *What matters most*. Wellington, New Zealand.

- Ministry of Education. (2004a). Using assessment tools. *Curriculum Update*, 54, April.
- Ministry of Education. (2004b). *Impact of teaching on variance in outcomes*. Wellington, New Zealand.
- Ministry of Education. (2005a). *ESOL Progress assessment guidelines*. Wellington, New Zealand: Learning Media.
- Ministry of Education. (2005b). *Key Competencies: The New Zealand Curriculum Project*. Wellington, New Zealand: Learning Media.
- Ministry of Education. (2005c). *Numeracy professional development projects*. Wellington, New Zealand.
- Ministry of Education. (2007a). TKI - *Curriculum Draft 2006-2007* Retrieved August, 16, from <http://www.tki.org.nz/r/nzcurriculum/>
- Ministry of Education. (2007b). *TKI Assessment Glossary*. Retrieved August, 13, 2007, from http.org./r/ars/glossary_e.php
- Ministry of Education. (2007c). *Te kete ipurangi: asTTle*. Retrieved August, 17, 2007, from http://www.tki.org.nz/r/asttle/index_e.php
- Ministry of Education. (2007d). TKI- Te Kete Ipurangi The Online Learning Centre. Retrieved August, 17, 2007 from <http://www.tki.org.nz/e/tki/>
- Ministry of Education. (2007e). *New Zealand Maths on line*. Retrieved August, 17, 2007 from <http://www.nzmaths.co.nz/numeracy/index.htm>
- Ministry of Education. (2007f). *New Zealand Maths on line: The Numeracy Project* Retrieved August, 17, 2007 from <http://www.nzmaths.co.nz/numeracy/index.htm>
- Ministry of Education. (2007g). Te Kete Ipurangi The Online Learning Centre: *Arts Page, Assessment Glossary*. Retrieved August, 17, 2007, from http://www.tki.org.nz/r/arts/glossary/index_e.php
- Ministry of Education & The University of Auckland. (2003). *Assessment Tools for Teaching and Learning*. Wellington, New Zealand: Ministry of Education.
- Ministry of Research, Science and Technology. (1998). *Mathematics in New Zealand: Past, present and future*. Wellington, New Zealand.
- Mitchell, F. W. (1968). *New Zealand education today*. Wellington, New Zealand: A.H. & A. W. Reed.
- Musial, D., Nieminen, G., Thomas, J., & Burke, K. (2009) *Foundations of meaningful education assessment*. New York: McGraw-Hill Higher Education.

- Narin, K., & Higgins, J. (2007). New Zealand's neo-liberal generation: Tracing discourses of economic (ir)rationality. *International Journal of Qualitative Studies in Education*, 20 (3), 261-282.
- Nash, R. (1983). Four charges against TOSCA. *New Zealand Journal of Education Studies*, 18 (2), 154-165.
- Neill, A. (2000). Diagnosing misconceptions in mathematics using the assessment resource bank. *SET Research information for teachers*, 2000 (1), 40-45.
- NEMP (2002). National Education Monitoring Project. Health and Physical Education assessment results 2002. Dunedin. Education Assessment Research Unit. University of Otago.
- NEMP. (2007). National Education Monitoring Project. Retrieved July 17, 2007, from http://nemp.otago.ac.nz/1_reports.htm
- Olsen, M., & Morris-Matthews, K. (Eds.). (1997). *Education policy in New Zealand: 1990s and beyond*. Palmerston North, New Zealand: Dunmore Press.
- O'Neil, A. M. (2004). Mapping the field: an introduction to curriculum politics in Aotearoa/New Zealand. In A.M. O'Neil, J. Clark, & R. Openshaw (Eds.), *Reshaping culture, knowledge and learning. Policy and content in the New Zealand Curriculum Framework*. (Vol. 1) (pp. 71-86). Palmerston North, New Zealand: Dunmore Press.
- O'Neil, A. M. (2005). Individualism, enterprise culture and curriculum policy. In J. Codd & K. Sullivan (Eds.), *Education policy directions in Aotearoa New Zealand*. (pp.71-86). Victoria, Australia: Thompson, Dunmore Press.
- O'Neil, A. M., Clark, J., & Openshaw, R. (Eds.). (2004). *Reshaping culture, knowledge and learning. Policy and content in the New Zealand Curriculum Framework*. Palmerston North, New Zealand: Dunmore Press.
- Organisation for Economic Co-operation and Development. (1983). *Review of national policies for education: New Zealand*. Paris: OECD.
- Organisation for Economic Co-operation and Development. (1995). *Measuring the quality of schools*. Paris: OECD.
- Parkin, C., Parkin, C., & Pool, B. (2002). *Probe reading assessment*. Lower Hutt, New Zealand: Triune.
- Patton, M, Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, Ca: Sage.
- Peddie, R. (2000). Evaluation of the assessment for better learning professional development programme. Retrieved December 12, 2007, from Numeracy

- Peters, M. L. (1979). *Diagnostic and remedial spelling manual*. London: McMillan Education.
- Peters, M., & Marshall, J. (1996). The politics of curriculum: Busnocratic rationality and enterprise culture, *Delta*, 48(1), 33-46.
- Phillips, D. (1998). National assessment policy in education. *New Zealand Annual Review of Education*, 8, 14-166.
- Phillips, G., McNaughton, S., & McDonald, S. (2001). *Picking up the Pace: Effective literacy interventions for accelerated progress over transition into decile 1 schools*. Wellington, New Zealand: The Child Literacy Foundation and Wolf Fisher Research Centre for the Ministry of Education.
- Picciano, A. G. (2006). *Data driven decision making for effective school leadership*. Upper Saddle River, NJ: Pearson.
- Pinder, M., Pinder, D., & Adams, G. (2001). *Wellsford School Mathematics Mastery Profiles Level 2*. Auckland, New Zealand: Pinda Publications.
- Poskitt, J. (2002). National consultation on the exemplars: What difference does it make. *SET Research information for teachers*, 2002(3).
- Poskitt, J., Brown, G., Goulton, F. & Taylor, K. (2002). *Researching development of the national exemplars: impact on teaching and learning*. Paper presented at NZARE Conference, Massey University, Palmerston North. December, 2002.
- Presseley, M., Wharton-McDonald, R., Allington, R., Black, C. C., Morrow, L., & Tracey, D. (2001). A study of effective first-grade literacy instruction. *Scientific Studies of Reading*, 5(1), 35-58.
- Reid, N., Croft, A. C., & Jackson, P. F. (1978). *Progressive Achievement Tests Study Skills Teachers' Manual*. Wellington, New Zealand: NZCER.
- Reid, N., & Gilmore, A. (1983). Pupil performance on TOSCA: Some additional information. *New Zealand Journal of Education Studies*, 18 (1), 13-31.
- Reid, N., & Gilmore, A. (1983). Test bias! Test bias? *SET Research information for teachers* 2, (item 16).
- Reid, N., Johnson, R. L., & Elley, W. (1994). *Progressive Tests of Achievement: Listening Comprehension - Teachers Manual*. Wellington, New Zealand: NZCER.
- Reid, N. A. (1977). PAT: Study Skills Test. *SET Research information for teachers*, 2, (item 3).

- Reid, N. A. (1993). *Progressive Achievement Test of Mathematics: Teachers Manual*. Wellington, New Zealand: NZCER.
- Reid, N. A., & Elley, W. B. (1991). *Progressive Achievement Tests of Reading (Teachers Manual)*. Wellington, New Zealand: NZCER.
- Renwick, M., & Gray, A. (1995). *Implementing the New Zealand curriculum in primary schools*. Wellington, New Zealand: NZCER.
- Robertson, S., Thrupp, M., Dale, R., Vaughan, K., & Jacka, S. (1997). A review of ERO. Final Report to the PPTA. Auckland, New Zealand: Auckland Uniservices Ltd, The University of Auckland.
- Robson, C. (2002). *Real world research* (2nd ed.). Malden, Ma: Blackwell.
- Ross, J. A. (2004). Effects of running record assessment on early literacy achievement. *Journal of Education Research*, 97 (4), 186-195.
- Rudman, H. C. (1995). Progressive Achievement Test of Reading [Electronic Version]. *The Mental Measurement Yearbook*, 12. Saunders, L. (1999). *Evaluation of curriculum reform in Solvenia: Approaches to evaluation*. Slough, UK: National Foundation for Educational Research.
- Sax, G. (1997). *Principles of educational and psychological measurement and evaluation*. Belmont, CA:Wadsworth.
- Scott, D., & Morrison, M. (2006). *Key ideas in education research*. Boston: Continuum.
- Shaklee, B. D., Barbour, N. E., Ambrose, R., & Hansford, S. J. (1997). *Designing and using portfolios*. Boston: Allyn and Bacon.
- Shuker, R. (1987). *The one best system? A revisionist history of state education in New Zealand*. Palmerston North: The Dunmore Press.
- Silverman, D. (2000). *Doing qualitative research a practical handbook*. Thousand Oaks, Ca: Sage.
- Simon, B. (1990). The Education Reform Act: Causative factors. In B. Broadfoot, P. Dockrell, C.Gipps, W. Harlen, & D. Nuttall (Eds.), *Policy issues in national assessment*. (pp 28-42) Clevedon, UK: Multilingual Matters Ltd.
- Skinner, B. F. (1953). *Science and human behaviour*. New York: Macmillan.
- Smith, A. B. (1982). School records: A critique on recent developments. *New Zealand Journal of Educational Studies*, 17(2), 180-184.
- Smith, J., & Elley, W. (1994). *Learning to read in New Zealand*. Auckland, New Zealand: Longman Paul.

- Smith, M. (Ed.). (2000). *Using Running Records: A resource for New Zealand classroom teachers*. Wellington, New Zealand: Learning Media for the Ministry of Education.
- Smith, N., & Nelley, E. (2002). *PM Benchmark: Kit 2 Teachers Notes*. Melbourne, Australia: Nelson Thompson Nelson.
- Snook, I., Adms, P., Adams, R., Clark, J., Codd, J., Collins, G., Harker, R., O'Neil, J., & Pearce, D. (1999). Education reform in New Zealand 1989-1999: Is there any evidence of success? *Delta*, 51 (1), 23-54.
- Stake, R. E. (2000). Case Studies. In N.K. Denzin, & Y. S. Lincoln, (Eds.), *The handbook of qualitative research* (2nd ed.) (pp. 435-454). Thousand Oaks, Ca: Sage. 435-454
- Stobart, G. (2006). The validity of formative assessment. In J. Gardner (Ed.), *Assessment and learning*. (133-146) London, UK: Sage.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks, Ca: Sage.
- Sullivan, K. (1992). The myth of partnership: Education reform and teacher disempowerment. *New Zealand Annual Review of Education*, 2, 151-165.
- Sullivan, K. (2002). Education policy and practice in contemporary Aotearoa-New Zealand. *McGill Journal of Education*, 37(1), 6-18.
- Sutton, R. (1998). *School-wide assessment. Improving teaching and learning*. Wellington, New Zealand: NZCER.
- Symes, I., & Timperley, H. (2003). Using achievement information to raise student achievement. *SET Research information for teachers*, 2003(1), 36-39.
- Teachers oppose compulsory primary school tests. (1998, May 8). *The Dominion*, p.3.
- Thomas, G., Tagg, A., Holton, D., & Brown, G. T. L. (2002). *Numeracy item signature study: A theoretical derived basis*. Auckland: University of Auckland, Project asTTle.
- Thomas, G., & Ward, J. (2001). *An evaluation of the count me in too pilot project. Exploring issues in mathematics education*. Wellington, New Zealand: Ministry of Education.
- Thorndike, R. M. (1997). *Measurement and evaluation in psychology and education*. New York: Merrill.

- Thrupp, M., Harrold, B., Mansell, H., & Hawksworth, L. (2000). *Mapping the cumulative impact of educational reform: A study of seven New Zealand schools*. Hamilton, New Zealand: University of Waikato.
- Thrupp, M. (1997). School mix and the outcomes of educational quasi-markets. In M. M. Olsen & K. Morris (Eds.), *Education policy in New Zealand: the 1990s and beyond*. (pp.372-389). Palmerston North: The Dunmore Press.
- Thrupp, M. (2004). Official school improvement in England and New Zealand: A cautionary comparison. *New Zealand Annual Review of Education*, 14, 39-58.
- Thrupp, M. (2005). Importing England's third way. In J. Codd & K. Sullivan (Eds.), *Education policy directions in Aotearoa New Zealand*. (pp.104-114). Victoria, Australia: Thomson, Dunmore Press.
- Timperley, H. (2003). Evidence-based leadership: the use of Running Records. *New Zealand Journal of Education Leadership*, 18, 65-74.
- Timperley, H. (2003). *Shifting the focus: Achievement information for professional learning*. Wellington, New Zealand: Ministry of Education.
- Timperley, H., & Parr, J. (2004). *Using evidence in teaching practice: Implications for professional learning*. Auckland, New Zealand: Hodder Moa Beckett.
- Timperley, H., Phillips, G., Wiseman, J., & Fung, I. (2003). *Shifting the focus: Achievement information for professional learning: A summary of the sustainability of professional development in literacy - Parts 1 and 2*. Wellington, New Zealand: Ministry of Education.
- Tomlinson, S. (2007). Ruthless assessment in a post-welfare in U.K. society. In C. Sleeter. (Ed.), *Facing accountability in education: Democracy and equity at risk*. (pp.172-187). New York: Teachers College Press.
- Torrance, H. (1993). Combining measurement-driven instruction with authentic assessment: Some observations of National Assessment in England and Wales. *Education Evaluation and Policy Analysis*, 15(1), 81-89.
- Torrance, H. (Ed.). (1995). *Evaluating authentic assessment*. Buckingham, UK: Open University Press.
- Treasury. (1990). *The Treasury briefing to the incoming government*. Wellington, New Zealand.
- Tuck, B. F. (1983). Education and tests of scholastic ability: Is there a baby in the bathwater? *New Zealand Journal of Educational Studies*, 18(2), 165-171.

- Visser, H., & Bennie, N. (1996). *An evaluation of the resource Beginning School Mathematics*. Wellington, New Zealand: Ministry of Education, Research Section.
- Ward, A. W., & Murray-Ward, M. (1999). *Assessment in the classroom*. Belmont, CA: Wadsworth .
- Weedan, P., Winter, J., & Broadfoot, P. (2002). *Assessment: What's in it for schools*. London: Routledge, Falmer, Taylor and Francis Group.
- Weiner, L. (2007). NCLB, Education and the world bank. In C. Sleeter. (Ed.), *Facing accountability in education: Democracy and equity at risk*. (pp. 94-117) New York: Teachers College Press.
- Whetton, C. (2004). *More or less? Current assessment issues in England*. Paper presented at the American Research Association.
- Whetton, C. (2004). *Reflections on fifteen years of national assessment*. Paper presented at the International Association for Educational Assessment.
- Williams, H. (1999). Assessment in primary schools. *New Zealand Journal of Educational Administration*, 14-20.
- Willis, D. (1992). Assessment trends in 1992: Future directions handicapped by a legacy from the past? *New Zealand Annual Review of Education*, 2, 245-262.
- Willis, D. (1992b). Educational assessment and accountability: A New Zealand case study. *Journal of Education Policy*, 7 (2), 202-221.
- Worthen , B. R. (1993). Critical issues that will determine the future of alternative assessment. *Phi Delta Kappan*, 80, 444-454.
- Wylie, C. (1997). *Self-managing schools seven years on: What have we learnt?* Wellington, New Zealand: NZCER.
- Wylie, C. (1999). *Ten years on: How schools view education reform*. Wellington, New Zealand: NZCER.
- Young, J. W. (2003). STAR Supplementary Tests of Achievement in Reading [Electronic Version]. *Mental Measurement Yearbook*, 15.

APPENDIX A: Principals and Senior Management Interview Form

Profile

Position

Time in present school

Years teaching and levels

Qualifications

Other relevant points

Introduction

I am interested in your school's assessment policies and practices – that includes all the formal assessment that takes place in the school such as PAT tests, Stars Reading tests, asTTle etc, as well as the informal assessment that teachers undertake in class on a day to day basis.

I am also interested in your reporting procedures including reporting to parents, the Ministry of Education's formal planning and reporting process, and reporting to the BOT.

I am particularly interested in the manageability and utility of assessment. Here is an outline of my understanding of the concepts of manageability:

Manageability

Refers to such things as:

- How much time it takes to carry out the assessment.
- How much time the assessment takes to mark.
- What is the cost of the materials and how difficult it is to organise them.
- Is the testing process easy to follow and how clear are the marking and administration guidelines?
- Do teachers need special training to administer and interpret the assessment?

Utility

Refers to how useful is the information to:

- Teachers
- The school management
- The BOT
- Students
- Parents
- Other agencies.

Questions

1. Could you tell me briefly about the key features of your school's assessment system? In a calendar year what assessment procedures would take place?
2. What is working well at present?
3. What do you see as challenging?
4. What do you think needs changing and why?
5. Has your assessment system/practices changed significantly in the last three years? If so how and why?
6. Are there any changes you see as being made in your assessment system in the next 6 months? Could you describe the changes you are considering and the rationale for them?
7. Within your school's assessment system who has responsibility for the different aspects of the assessment procedures and their implementation? – Who does what?
8. At present are there any big gaps between your thinking about assessment practices and other staff members? If so who did it concern and what were the issues? Have there been any significant gaps in the past?
9. Can you tell me about your assessment background – how did you learn what you know?

Have you attended any courses – when – what duration?

Is there significant reading that influenced you?

Has your school done ATOL?

Has anybody on your staff attended any assessment courses?

Have any people significantly influenced your thinking on assessment? If so, who?

10. Could you summarise for me what you think are the most important things schools need to have in place in terms of assessment?

For the questions 11-18 about specific assessment tools, I want you to distinguish between their manageability and their utility.

11. Is your school using asTTle – what for? When did you start? What have you found out? How manageable have you found it? How would you describe its utility?
12. Is your school as a whole or teachers within the school, using The Exemplars? If so how are they being used? How manageable have you found them? How would you describe their utility?
13. Is your school as a whole, or teachers within the school, using the Assessment Resource Banks? If so how are they being used? How manageable have you found them? How would you describe their utility?
14. Does the school use PAT Tests. If so which tests are administered and how is the information used? How manageable have you found them? How would you describe their utility?
15. Is the school using STAR? If so when was it introduced? How is it used in the school? How manageable have you found it? How would you describe its utility?
16. Could you describe how Running Records are used in your school? How manageable do you found them? How would you describe their utility?
17. Does you school carry out Numeracy Project Assessments? If so what form do they take? How manageable have you found them? How would you describe their utility?
18. Do you use the NEMP assessment tools at all? If so how are they used? Has any NEMP material/information influenced school practice? If so how?

.....

19. Does the school operate a system of student profiles/ or portfolios. How does this operate? Have they changed in the last three years, if so how? How manageable are they? How would you describe their utility?
20. Most teachers and principals believe observation is an important part of assessment? What are your thoughts on this? Is there some formal element of observation in your school procedures?
21. What are your assessment procedures in the areas of?
- Social Studies
 - Science
 - The Arts
 - Te Reo
 - P.E.
22. Could you tell me about New Entrant assessment procedures in your school? How manageable have you found them? How would you describe their utility?
23. Does your school have assessment procedures for the Essential Skills? If so, how do these operate?
24. Have you given any thought to implementing and assessing the key competencies?
25. Does the school have policy and procedures to assess higher order thinking? If assessing higher order thinking, how is it done?
26. Tell me about how you implement formative assessment in your school.
27. How have you implemented the Ministry's planning and reporting requirements? Has this been of value? If so, how?
28. Could you describe your system for reporting to the BOT. Has it changed in the past three years? Are there any challenges associated with this?
29. What school management system have you chosen? How do you think this will change assessment and reporting practice in the school?
30. Could you describe your system of reporting to parents? Has that changed over time? What is working well? Are there challenging aspects or elements you think you would change?

31. What is your understanding of the term *evidence-based teaching*? Can you give me examples of it in your school?
32. Has ERO influenced your assessment practices? If so, how?
33. Overall do you think what is asked of schools in terms of assessment is manageable and useful?
34. Do you have any other comments to make regarding the manageability and utility of assessment in your school?

APPENDIX B: Teacher Interviews

Profile

Position

Time in present school

Years teaching and levels

Qualifications

Other relevant points

Introduction

I am interested in both your classroom and whole school's assessment policies and practices – that includes all the formal assessment that takes place in the school including the likes of PAT tests, Star, asTTle etc, as well as the informal assessment that you undertake in class on a day to day basis.

I am also interested in the school's reporting procedures including: reporting to parents, the Ministry of Education's formal planning and reporting process, and reporting to the BOT.

I am particularly interested in the manageability and utility of assessment.

Manageability

Refers to such things as:

- How much time it takes to carry out the assessment.
- How much time the assessment takes to mark.
- What is the cost of the materials and how difficult it is to organise them.
- Is the testing process easy to follow and how clear are the marking and administration guidelines?
- Do teachers need special training to administer and interpret the assessment?

Utility

Refers to how useful is the information to:

- Teachers
- The school management
- The BOT
- Students
- Parents
- Other agencies.

Questions

1. Could you tell me briefly about the key features of your school's assessment system? In a calendar year, what assessment procedures would you see taking place?
2. What is working well at present?
3. What do you see as challenging?
4. What do you think needs changing and why?
5. Has your assessment system/practices changed significantly in the last three years? How and why?
6. Are there any changes you are aware of being made in your assessment system in the next 6 months?
7. Within your assessment system, who has responsibility for the different aspects of the assessment system and its implementation? Who does what?
8. At present are there, or have there been in the past, any big gaps between your thinking about assessment practices and other staff members especially the senior management? If so, what are, or have been, the issues?
9. What do you find most rewarding about assessment?
10. What do you find most frustrating about assessment?
11. Can you tell me about your assessment background – how did you learn what you know?

Have you attended any courses – when – what duration?

Has there been any significant reading that has influenced your views?

Has your school done ATOL?

Have any people significantly influenced your thinking on assessment? If so who?

For the following questions 12-19 about specific assessment tools I want you to distinguish between their manageability and their utility.

12. Is your school using asTTle – what for? , When did you start? What have you found out?
13. Is your school as a whole or teachers within the school, using The Exemplars? If so how?
14. Is your school as a whole, or teachers within the school, using the Assessment Resource Banks? If so how are they being used? How manageable have you found them? How would you describe their utility?
15. Does the school use PAT Tests. If so how and what for? How manageable would you describe them? How would you describe their utility?
16. Is the school using STAR? If so when was it introduced? How is it used in the school? How manageable would you it? How would you describe its utility?
17. Could you describe how Running Records are used in your school? How manageable would you describe them? How would you describe their utility?
18. Does you school carry out Numeracy Project Assessment? If so what form does it take? How manageable would you describe them? How would you describe their utility?
19. Do you use the NEMP assessment tools at all? If so how? Has any NEMP material/information influenced your or the schools practice if so how?

.....

20. Does the school operate a system of student profiles or portfolios? How does this operate? Have the changed in the last three years if so how? How do you personally use this information? What value would you put on the system? How manageable would you describe them? How would you describe their utility?
21. Most teachers and principals believe observation is an important part of assessment. What are your thoughts on this? Is there some formal element of observation in your school procedures?

22. What are your assessment procedures in the areas of?
- Social Studies
 - Science
 - The Arts
 - Te Reo
 - P.E.
23. Does your school have assessment procedures for the Essential Skills? – How do these operate?
24. Has the school given any thought to implementing and assessing the key competencies?
25. Does the school assess higher order thinking? If so how?
26. Tell me about formative assessment in your school and in your classroom?
27. What has the School done in terms the Ministry of Education's *planning and reporting/target setting*? How has this affected you as a classroom teacher?
28. Could you describe your system of reporting to parents? Has this affected changed over time? What is working well? Are there challenging aspects or elements you think you would change?
29. What is your understanding of the term *evidence-based teaching*? Can you give me examples of it in your school?
30. Has ERO influenced your own or the schools assessment practices? If so how?
31. In addition to the school and syndicate procedures do have any assessment practices such as testing and recording, observations –or just making mental notes as you go, that are just yours and specific to your own individual classroom practice. Could you describe them for me?
32. Overall do you think what is asked of you as a teacher in terms of assessment is manageable and useful?
33. Do you have any further comments regarding the manageability and utility of assessment?

APPENDIX C: Consent Form - Principal

PROJECT TITLE: The Manageability of Assessment in New Zealand Primary Schools 1993 –2006: Case Study of Three Schools

Name of participating School:

Name of investigator: John Young

- (1) I consent to the participation of my school in the project named above. Details of the project and my school's participation, including my own, have been explained to my satisfaction.
- (2) I acknowledge that:
 - (a) I have been informed that the School or any individual participant is free to withdraw from the project at any time without explanation or prejudice, and to withdraw any data supplied.
 - (b) The project is for the purpose of research.
 - (c) I have been informed that the anonymity of the School and any individual associated with the school will be protected.
- (3) I understand that my interview will be audio-taped and a copy of the transcript will be returned to me for verification.
- (4) I understand that the audio tapes will be wiped two years after the publication of the research.
- (5) The school and any individual associated with the School will be referred to by pseudonym and not be identifiable.
- (6) At the completion of the research, a summary will be sent to both the school and myself.

Signature School Principal

Date

APPENDIX D: Consent Form - Staff Member

PROJECT TITLE: The Manageability of Assessment in New Zealand Primary Schools 1993 –2006:
Case Study of Three Schools

Name of participating School:

Name of investigator: John Young

- (1) I consent to participate in the project named above. Details of the project and my participation have been explained to my satisfaction.
- (2) I acknowledge that:
 - (b) I have been informed that I am free to withdraw from the project at any time without explanation or prejudice, and to withdraw any data previously supplied.
 - (b) The project is for the purpose of research.
 - (c) I have been informed that my anonymity, and the anonymity of the School, will be protected.
- (7) I understand that my interview will be audio-taped and a copy of the transcript will be returned to me for verification.
- (8) The school and any individual associated with the School will be referred to by pseudonym.
- (9) At the completion of the research, a summary will be sent to both the school and me.

Signature:

Date

APPENDIX E: Letter to Principals/Staff Member

I am writing to ask if you are willing to participate in a research project related to my doctoral studies at Victoria University of Wellington (VUW), School of Education Studies. I wish to interview you in relation to your experiences at [name of school]. I have obtained the permission of the BOT of [name of school] to interview you.

The study aims to examine assessment from 1993 to 2006 through a case study of three New Zealand primary schools. The particular area of focus is how individual schools developed procedures to meet the changes in assessment requirements introduced as part of the educational reforms of the early nineties. The study will trace how the schools in the study have adapted to changes in government policy, as well as the introduction of new assessment tools, over the last thirteen years.

The study involves interviews with principals, other school management staff, some teachers, BOT members and, where necessary, former staff and BOT members. Interviews with the latter will be to fill in gaps due to turnover of school staff and BOT members. The research will also include examination of the school's assessment documentation such as policies, procedures and reports to the BOT.

Interviews with former staff and BOT members are expected to take no more than 30 minutes each. Interviews will be taped and transcribed. The transcription of your interview will be returned to you for verification. Two years after the completion of the research the tapes will be wiped. Any information you provide will be kept confidential to the researcher, the supervisor, and the person who transcribes the tape recording of our interview.

The anonymity of the School, as well as individuals associated with the School, will be protected. Details will be kept secure in a separate, password-protected computer file. In the final report, the School will be referred to by a pseudonym. I will remove references to personal information that might allow someone to identify you or any other individual associated with the school. The data will be kept for one year from the date of publication, before being destroyed.

The School's participation in this study, as well as your own contribution, is completely voluntary. Should you wish to withdraw any data you have supplied before analysis is complete, you are free to do so without having to give reasons or without penalty of any sort.

The study has received approval from the VUW Human Ethics Committee (College of Education Subcommittee).

If you are willing to participate, please indicate that you have read and understood this information by signing the accompanying consent form and returning it in the envelope provided.

If you would like further information, please contact either the research supervisor at VUW (Professor Cedric Hall: cedric.hall@vuw.ac.nz; phone 04-463 9772) or myself directly (John Young: principal@stjosephsuh.school.nz; phone 04-428 4910).

Yours sincerely,

John Young

APPENDIX F: Focus Group Process

Purpose

1. Test the credibility of Three Waves of Assessment and Four Phases of Assessment Change frameworks with a group of principals.
2. Get a fuller understanding of factors that influence assessment change in schools.
3. Check summary of the assessment practice prior to the reforms.

Members

- Current principals
- Had been teaching or held principal positions prior to 1990
- Have been principals or APs/DPs since 1995

Process

The researcher will give a short presentation on:

- Three waves of assessment
- Four phases of assessment change
- Factors that influence assessment change: School leaders, professional development, Networks with other school (informal and formal), MoE, ERO

The group will be given a copy of section 3.2: Assessment Activity in New Zealand Primary Schools in the 1980s (pp.38-410).

The researcher will then leave the room and the facilitator will then lead the process.

There will be poster of the four major themes in the room:

- Three waves of assessment
- Four phases of assessment change
- Factors that influence assessment change: School leaders, professional development, Networks with other school (informal and formal), MoE, ERO
- Assessment Activity in New Zealand Primary Schools in the 1980s

The group will be asked to reflect individually on how what has been presented matches with their understanding and write comments on post-it notes and attach them to the posters.

The facilitator will then lead discussion on the four themes. A summary will be written by the facilitator and checked by the group.

The researcher will summarise the discussion and notes and send a copy to each group member.

