

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Rethinking Assessment:

The Challenging Issues for Schools and Teachers

A thesis submitted as partial fulfillment
of the requirements for the degree of
Master of Educational Administration

Massey University: Albany

Jan Hill

1998

THESIS MEMORANDUM

Private Bag 11 222,
Palmerston North,
New Zealand
Telephone: 64 6 356 9099

TO: Helen Sneddon
Learning & Teaching - HOKOWHITU

FROM: RECORDS OFFICE

DATE: 08 June 1999

SUBJECT: RETURN OF THESIS FOR LIBRARY PURPOSES

x 8961 *Wayne 8/6/99* *Edward's*

Can a copy of Jan Hill's thesis be returned to the Records Office, Level 4, Registry (Turitea), as the student has now been recommended for the degree of M.Ed.Admin.

Thanking-you

*Have been unable to ~~locate~~
Obtain second copy.*



Massey University Library Thesis Copyright Form

Title of thesis:

- (1) (a) I give permission for my thesis to be made available to readers in Massey University Library under conditions determined by the Librarian.
- (b) I do not wish my thesis to be made available to readers without my written consent for ... months.
- (2) (a) I agree that my thesis, or a copy, may be sent to another institution under conditions determined by the Librarian.
- (b) I do not wish my thesis, or a copy, to be sent to another institution without my written consent for ... months.
- (3) (a) I agree that my thesis may be copied for Library use.
- (b) I do not wish my thesis to be copied for Library use for ... months.

Signed



Date

The copyright of this thesis belongs to the author. Readers must sign their name in the space below to show that they recognise this. They are asked to add their permanent address.

NAME AND ADDRESS

DATE 4.2.98

Jan Hill
2/22 Norwood Rd
Bayswater
North Shore, Auckland

Jan HILL

2/22 Norwood Rd

Bayswater

Auckland

Ph 4456214

WORK

IPDER Building 53
Albany Campus
4439388

Supervisor:

Dr Tenny Postcitt

Dept. of Learning and Teaching

Massey University: Palmerston North

Ph. X 8959

DECLARATION

I declare that this thesis represents my own work, except where due acknowledgment is made, and it has not been previously included in a thesis, dissertation or report submitted to this University or any other institution for a degree, diploma or any other qualification.

A handwritten signature in black ink, appearing to read 'Jan Hill', is written over a horizontal dashed line.

Jan Hill

Abstract

This thesis examines how four urban primary schools used changes to their assessment practices as a means to improve the learning opportunities and outcomes of their students. In 1993, a new national curriculum was introduced into New Zealand schools and this was accompanied by legislation and guidelines mandating new requirements in assessment. These reforms were occurring against a backdrop of burgeoning developments in assessment internationally. The study documents how the four schools responded to the new demands, to the challenges posed by alternative approaches to assessment and how they were able to incorporate and build on their previous assessment practices.

Action research was selected as the methodology and was used both by the schools and the researcher on two distinct yet overlapping levels. It gave the schools the opportunity to take ownership of the issues that emerged and greater control over the research process. The data gathering strategies woven into the programme included group discussions, individual interviews, observations, questionnaires, document analysis and a Help Desk which formalised contact with the schools between school-based visits and workshops and provided the researcher with valuable ongoing insights into the work of the schools. The thesis incorporates an analysis of methodological issues relating to collaboration, the tension between first and second order domains of action research and difficulties the schools experienced in relation to the action research process itself.

The data revealed a number of emerging themes. Summative rather than formative assessment practices dominated the aspects of assessment the schools selected to work on as a result of their baseline data collection. All of the schools put new school-wide recording systems in place. This had the most impact on the practices of the teachers who, in the past, had assessed in an ad hoc way, often based on 'gut reaction'. For many teachers, it was first time they had investigated and understood assessment theory. The data also highlighted that many do not have a level of technical assessment expertise that allows them to design basic assessment tasks that they know to be valid measures or to analyse and use the results with confidence. Although the common orientation of the teachers was towards a learner-centred philosophy, there is a considerable gap between theory and practice. Rethinking assessment practices provided some of the teachers with a vehicle for shifting the locus of control in their classrooms. Traditional reporting practices were also challenged and possibilities emerged for new assessment partnerships with parents. For some teachers, giving honest feedback to students and parents, in particular, poses a dilemma. The transfer of assessment information both within and between schools remains a possible barrier to student achievement. The thesis concludes that, overall, the curriculum and assessment changes appear to have had a positive impact on the learning and teaching programmes in the four schools, but that much professional development work remains to be done if both contemporary assessment theory and policy are to translate into classroom practice.

Acknowledgments

There are many people who have contributed significantly to this research even though my name appears on the front cover. I owe a particular debt to the principals and staff of the four schools that participated in the research, particularly the members of the four Lead Teams. The research was dependent on their participation and the outcomes of the research are a tribute to their expertise, their willingness to share that expertise and their preparedness to take risks. The introduction of the changes outlined in this document demanded a great deal of them and without their efforts and determination to do the best for their schools and their students, none of these changes would have happened.

Many other people have contributed both formally and informally to this thesis in many different ways. I acknowledge the invaluable contributions of both Dr Jenny Poskitt and Dr Ken Ryba, especially Jenny, who did her best to keep me on task and constantly challenged me to go the extra mile. I owe special thanks to my friend and colleague Kay Hawk, for her encouragement and understanding when the going got tough, as it did from time to time. Also, to Norma Roberts for her secretarial support and for helping out in many, many practical ways. And to Julie who lent me her spare room for some weeks so that I could escape to write.

Lastly, and most important of all, my love and gratitude to Noel and Sam, for all those weekends without me, for often doing my share of the tasks, for understanding how important this study was to me and without whom I could never have achieved this dream.

Table of Contents

ABSTRACT	ii
ACKNOWLEDGEMENTS	iii
CHAPTER ONE: Introduction	
Introduction	1
The context: assessment	2
The research problem	3
The research questions	3
The process: action research	5
My role in the research	6
Organisation of the thesis	7
CHAPTER TWO: Background to the study	
Introduction	9
The era of the Proficiency Examination	9
Register of Progress and Achievement	11
Primary Progress Records and educative assessment	14
The impact of the New Right, new legislation and the new curriculum	17
Summary	21
CHAPTER THREE: Literature Review	
Introduction	23
The interaction of assessment with theories of learning	24
The role of assessment	27
Approaches to assessment	31
Designing assessment tasks and questions of validity	35
The role of teachers in the assessment process	38

Reporting to parents	41
Summary	44

CHAPTER FOUR: The methodology in theory

Introduction	46
The evolution of action research	46
The key principles that underpin action research	49
Action research models	53
Action research tools	56
Summary	58

CHAPTER FIVE: The methodology in practice

Introduction	60
Origins of the action research process in the four ABLE schools	61
The action research model	65
Tools used to gather the data	67
How action research principles were reflected in the study	70
Data analysis	73
Ethical issues	74
Methodological limitations	79
Summary	81

CHAPTER 6: Emergent Themes

Introduction	82
Understanding assessment theory	83
Understanding the curriculum and the expectations of Ministry	88
Planning and design	93
Control	98
Honesty	105
Consistency	108
Summary	112

CHAPTER 7: Insights and conclusions

Introduction	114
Autonomy and accountability	115
Assessment literacy	118
Assessment partnerships	121
Methodological insights	125
Future research	129
Summary	130

BIBLIOGRAPHY	132
---------------------	-----

APPENDICES

- A. National Administration Guidelines
- B. Letter of introduction to the Contract and Expression of Interest form
- C. Baseline post-box question questionnaire
- D. Sample action plans
- E. Sample of fax form used to set up the workshop agendas
- F. End-of-study lead team evaluations used to review the year's work and to plan for the following year
- G. Samples of annotations used to code and analyse transcriptions and field notes; a sample analytic memo

LIST OF TABLES

1. Statistics for each school in relation to the selection criteria
2. The action research model developed for the study
3. The list of tools used to gather the data, when they were used and how the data were recorded
4. Procedures used in the study to reflect the principles of action research
5. The key purposes of assessment as identified by the four study schools
6. Techniques the teachers are knowledgeable about and are confident in using

CHAPTER ONE

INTRODUCTION

'Assessment is one of the most powerful levers for enhancing student learning and bringing about positive educational change' (Earl and LeMahieu, 1997: 158).

Introduction

Rethinking Assessment: the challenging issues for schools and teachers is based on the study of four primary schools who participated in a year-long assessment contract funded by the Ministry of Education¹. There were a number of similar contracts operating across the country at the time, all with the purpose of improving the assessment systems and practices of the schools and the classroom practitioners. Internationally, assessment is becoming a very 'high stakes' activity for students, teachers and schools. At the same time, there is widespread dissatisfaction with the available assessment procedures. Assessment reform, like other educational changes, is going through a period of rapid change and uncertainty. In New Zealand, these assessment changes are occurring in tandem with the introduction of the new national curriculum documents. Because of the well-established importance of ownership in any change process (Dalin and Rolff, 1993: 14) and because the schools had a myriad of aspects of assessment from which to select a focus for development, action research was considered to be the most appropriate research methodology. It was intended that the schools would identify and clarify the issues of importance to each of them and that subsequent action research cycles would allow them to explore these in-depth and to connect new layers of meaning to them. Therefore, this thesis investigates not only the content or substance of the research (assessment) but also the impact of the process (action research) on the changes that occur.

¹ From now on in the text, the Ministry of Education will be referred to as the Ministry.

The content: assessment

Historically, assessment practices in New Zealand primary schools have been dominated by a psychometric approach to educational measurement, a culture of testing and a focus on norm-referencing as a means of identifying students' levels of achievement (O'Neill, 1994). These approaches were characterised by the Proficiency Examinations, which shaped the primary curriculum until their abolition in 1937, and the Register of Progress and Achievement and the Primary School Record Cards that were the hallmarks of assessment practices through to the late 1980s. While teachers were expected to grade each student on a norm-referenced, 1-5 scale, how that norm referencing was decided was left entirely over to the schools. Schools and teachers also had considerable autonomy to make decisions about any other assessment information that was recorded in the Register. In 1989, these documents were replaced by the Primary Progress Records which reflected a move away from norm-referencing and towards, what Blackmore (1988:46) refers to as, a more educative assessment paradigm.

Educative assessment practices, based on constructivist theories of learning (Broadfoot, 1998) emphasise the central importance of the learner in the process and are more concerned with measuring a student's progress against previous performances than against the performances of their peers. On that basis, credence is given to wide range of methods of assessment and both quantitative and qualitative data are used to develop an holistic view of a student's progress. The Primary Progress Records were, primarily, qualitative documents. While teachers made professional judgements about what they recorded and therefore what counted, in their view, as measures of achievement, the system of recording was still being mandated by the Ministry.

The 1990s heralded a shift towards greater accountability of schools for the learning outcomes of students. A new national curriculum was prescribed for all schools across the country and accompanying legislation (Ministry of

Education, 1993) indicated that assessment would become the framework for controlling the delivery of the curriculum (Hill, 1997).

This study took place three years after the curriculum and the legislation were introduced. Schools were still coming to terms with the developments and, once it was publicised that schools were not bound to using the Primary Progress Records, an increasing number of them were experimenting with alternative systems. This was occurring against a backdrop of burgeoning developments in assessment. Research was highlighting the significant gains to learning of formative assessment and a whole new vocabulary, developing around alternative approaches to assessment, was becoming common currency. This included concepts such as performance assessment, self and peer assessment, standards-based assessment and the formulation of rubrics. Against this backdrop, the research problem and the questions that would provide a framework for the study were formulated.

The research problem

Because the reforms of 1993 were so new, there was little research that documented how schools were responding to requirements or how teachers were incorporating and building on previous assessment practices. The aim of the research was to document the challenging issues for schools and teachers as they came to terms with the new curriculum and implemented the Ministry's new assessment requirements and guidelines.

The research questions

The following questions were formulated to act as a guide to the research rather than to constrain or control it in any way. They were designed to be sufficiently broad as to accommodate the problems and issues identified by the four individual schools.

- What understanding of assessment do the teachers have before they begin the development process?
- What systems and practices do the schools have in place at the time the study begins?
- What changes are made to school-wide assessment systems and practices over the course of the year?
- Does the teachers' thinking about assessment change over the course of the year? In what ways?
- What do the schools consider to be the successes? Why are they considered to be successes?
- What are the key influences in bringing about any changes in their thinking and practices?
- What are the issues that emerge for both schools and teachers? Are they resolved? How? If not, why not?
- What differences are there, if any, in the assessment priorities of the Ministry versus those identified by the schools?
- What are they and how do schools address them?
- How does action research contribute to the development process? What are the successes? What are the difficulties?

Quite clearly, the questions called for a qualitative approach that would allow the issues to emerge from the study rather than beginning with a set of assumptions about the participants' perceptions, needs and approaches they should take to bringing about any changes.

The process: action research

Despite the continuing criticisms of practitioner inquiry as a legitimate form of educational research, there has also been growing support for its knowledge-generating potential (Zeichner and Noffke, 1998), its potential to actively engage teachers in the change process (Elliot, 1991) and the support it provides teachers in coping with the challenges and problems of implementing innovations (Altrichter, Posch and Somekh, 1993). Nonetheless, Johnston (1994) argues that it is not a natural process for teachers. They struggle with the unfamiliar, but basic, techniques of research. The demands of their jobs make it difficult for them to find the time to do the research and when they do so, their attention is drawn away from their main task of educating students. As a result, the outcomes are often deemed to be less rigorous and not up to an acceptable standard, even though action research uses the same methods as regular educational research. There is also evidence that even those who acknowledge practitioner research as a form of educational scholarship, have sometimes viewed it as an inferior form of research (Zeichner and Noffke, *ibid*). For their part, schools traditionally value craft knowledge of teachers above academic research and place a higher premium on the practical rather than the theoretical realm (Poskitt, 1994). These negative observations were not sufficient to deter me from employing an action research methodology. The motivation lay in the value I had learned to place in formative research.

In 1998/1989 I was employed by the Ministry as a 'developer', or facilitator, on a major project, the purpose of which was to explore ways schools could develop partnerships with their communities. Known as the Curriculum Review Exploratory Study (CRES), there was a research strand to the programme, which entailed each of the four developers being shadowed by a researcher for the eighteen months over which the project operated. Up until that point, I was woefully ignorant of research processes and dismissive of the world of the education academic. I had little idea of how research could benefit the practitioner. Because I had been selected for the position on the basis of my practitioner skills and my organisational ability, I also had to learn to facilitate,

in schools where many had been coerced into the programme. It was a vertical learning curve but one of the most exciting and rewarding in terms of my own personal and professional growth. A critical influence on that growth, and that of the schools, was the feedback given by the researchers and the critical reflection and action that resulted. For those individuals and schools that embraced the opportunities offered by the research, the outcomes were profound. I hoped that this research project was one way I could excite and enthuse other teachers, in the same way I had been, about the value of research and of being directly involved as a practitioner researcher.

Action research was a way of marrying the two worlds of research and teaching. The teachers would benefit from deeper and improved understandings of their craft theories and teaching practices and, as the academic, I would learn from applying the theory in context. The process would be reciprocal and collaborative.

My role in the research

Because the research was conducted alongside and was part of a development programme, the methodology also had to accommodate my role as the Contract facilitator. The research design needed to allow me to fulfil that role without compromising or directing the research outcomes. In order for those outcomes to reflect the schools' issues rather than mine, it was imperative that the research be controlled by the participants and informed by the group's ongoing reflection on outcomes. The context was ideally suited to action research and a commitment to the underlying principles of naturalistic inquiry. Action research necessitates that the research be of direct benefit to the four schools and the teachers who work in them. I wanted them to be active participants and decision-makers in the research process and to experience the value of research for informing their own practice.

There were two distinct and overlapping levels to this research. The first level (or first-order domain) concerned the ongoing work of the schools and the

Contract facilitator using an action research model. It involved each school group working with their own data as well as with the material I collected as the primary researcher. Using most of the activities associated with action research, the aim was to allow the schools to make positive and ongoing changes to their assessment systems and practices in each school as well as to help the lead teams² identify processes and strategies that, in their particular contexts, would be conducive to those changes being made.

The second level (or second-order domain) was my own work as the primary researcher where I was collecting data to better understand the current issues in assessment practice and to write a thesis. I used a range of qualitative techniques to collect the data: open-ended baseline surveys, evaluations, semi-structured interviews, participant observations and document analysis. As the data were gathered, it was analysed using the principles underlying grounded theory. Emerging themes were identified and more data collected to either reject or validate the continued inclusion of each theme. Although these two levels have been described as two distinct and separate entities, in practice, they were completely interwoven and inseparable.

Organisation of the thesis

The next chapter (chapter two) outlines the background to the study and describes the educational context in which the research took place. It provides an historical perspective of assessment practices in New Zealand primary schools and, in particular, traces the development of the summative practices that pervaded the delivery of the curriculum. Chapter three is a review of the literature on assessment. It examines the interaction of assessment with theories of learning, approaches to assessment and the role of the teacher and of other key stakeholders in the assessment process. The review highlights the lack of a substantial body of research into both summative and formative assessment practices in New Zealand primary schools. The following two

² The lead teams coordinated the development process in each school. Each team generally comprised the Principal, another senior staff member and one or two other teachers.

chapters (four and five) focus on the theory and application of action research. Particular attention is paid to the ways in which the principles of action research were reflected in the study. Finally, in chapters six and seven, the themes that emerged from the action research cycles are described and then discussed in light of the expectations of Ministry policy and in terms of what the literature suggests is good assessment practice. Chapter seven also evaluates the impact of the action research methodology on the development process in the schools. And so, the journey of the research begins.

CHAPTER TWO

BACKGROUND TO THE STUDY

We shall only teach better if we learn intelligently from the experience of shortfall; both in our grasp of the knowledge we offer and our knowledge of how to offer it. That is the case for research as the basis for teaching' (Stenhouse, 1983: 193).

Introduction

Assessment practices in New Zealand have changed markedly over the years. Developments in this country have been influenced by education research and changes in assessment practice in other Western countries. They have also been a reflection of the political ideologies and policies of their time. This chapter traces the history of assessment practice in New Zealand schools from the passing of the 1877 Education Act, when primary education became free, secular and (almost) compulsory, through to the introduction of the New Zealand Curriculum Framework and the National Education Guidelines in 1993. I have divided the historical background into what I see as three distinct periods, each characterised by significant changes to the national assessment system. The reasons for these changes are described and the key principles underlying the approaches are discussed. Accordingly, the chapter has been divided into four sections:

1. The era of the Proficiency Examination
2. The Register of Progress and Achievement
3. Primary Progress Records and a shift to educative assessment
4. The impact of the New Right, new legislation and the new curriculum

The era of the Proficiency Examination

Traditionally, assessment in schools in Western countries, including New Zealand, has been dominated by what Wolf *et al* (1991) refer to as 'culture of

testing', what Broadfoot (1992) calls an 'outmoded assessment culture' and Blackmore (1998), a 'technicist approach'. The roots of this technicist paradigm lie in developments that were taking place at the end of the nineteenth century and the early part of the twentieth century. Based on the work of psychologists like Fechner and Wundt in Germany and Galton in England, attempts were being made to "create a science of the social" (Eisner, 1993: 219). The intent was to apply methods that had been applied to nature, to the study of human beings. If humans were a part of nature and if nature could be understood, why not humans (*ibid*)? A combination of assumptions about the nature of ability and the mechanics of measurement were contributing to the development of a test and measurement movement in education (Willis, 1994: 163).

This technicist or psychometrics approach to assessment was characterised by a formal testing and examination regime that Torrance (1988: 7) described as totally divorced from the educational process and setting to which they were meant to relate. The emphasis was on standardisation and reliability and results were generally presented as grades or marks (Broadfoot, in Willis, 1994: 168). The focus was not on diagnosis of individual needs and measuring individual progress against past performance but on comparability for the purposes of ranking and sorting students.

O'Neill (1996: 3) describes the education system in New Zealand, at the time, as being well known for its academic, formalised, hierarchical and exam-driven nature. This was characterised by the introduction of the Proficiency Examination in 1899 which was sat by students at the end of their primary schooling³. Its initial purposes were to provide students with an end-of-primary school qualification and to achieve greater uniformity of standards and assessment procedures across schools. The latter was quickly overtaken by a need to satisfy a growing demand for access to free post-primary schooling and later, to determine who would qualify to enter the different 'grades' of high

³ Primary schooling concluded at the end of Standard 6, equivalent to the present day Year 8.

school - agricultural (district high schools), technical (technical high schools) and academic (secondary schools) (Openshaw et al, 1993, 197). The Proficiency Examination fitted neatly into the psychometric paradigm.

As a selective entrance test, the Proficiency Examination had a profound influence on the curriculum. Syllabuses were primarily shaped by the test prescriptions, schools and teachers were judged by the results of their students and the students themselves were sorted and classified according to their individual results. As reported in 1931 by the President of NZEI⁴ at the time:

There is still the Proficiency Examination casting its baneful shadow down on the school, and detrimentally affecting school methods and school life. So much has the examination method engrained itself into our school life; both primary and secondary, that not only do many teachers find great difficulty freeing themselves from its clutches but many parents have come to look on the annual full-dress examination as the only bona fide test and guarantee of a child's progress. The examination has to be passed and the battle won. If the child is successful, his [sic] is the glory and victory; if not, defeat with consequent discredit (ibid, 200).

Register of Progress and Achievement

The election of the first Labour Government in 1935 signalled a change in direction for the New Zealand education system. In 1937, the Proficiency Examination was abolished and, with the appointment of Dr C. E. Beeby in 1938 as the Assistant Director General of Education, a fifteen year overhaul of the entire curriculum began. More resources were directed into education and teachers were encouraged to interpret the curriculum more widely and creatively. In O'Neill's words (ibid), the activities and experiences of pupils were emphasised as was a decrease in the amount of time spent on grammar,

⁴ The New Zealand Educational Institute (NZEI) is the primary teachers' union.

spelling, oral reading and formal arithmetic and an increase in the time spent on music, art and crafts, nature study and physical education.

The development of thinking on assessment from this period of the 1940's and 50's through to the late 1980's, is reflected in the Register of Progress and Achievement and the Primary School Record Card. All teachers were required to keep a Register for each class each year. At the end of the year, the information in the Register was used to record achievement levels on individual student Record Cards. These Record Cards were a cumulative record of a child's progress and it was mandatory for them to be passed on to the next teacher or school.

The technician paradigm that was the basis for assessment practices up to this time was still dominant. For example, the notes on the use of the register, printed on the front and back covers, provided detailed guidance on the norm-referencing procedures to be followed.

When a teacher conducts a survey, the results should be assessed on a five-point scaleIn Part 1, achievement in the various subjects will be assessed on a five-point scale: 1,2,3,4,5. The five categories will be interpreted as meaning that, from the teacher's general experience of the children of the age concerned, he (sic) would expect that, out of a group of one hundred such pupils: five would receive the top rating of 1; twenty would receive the rating of 2; fifty, or about half would receive a rating of 3; twenty would receive a rating of 4; and five would receive the bottom rating of 5.

Rather than an examination to determine comparability, teachers were being asked to make professional judgements based on their experience of how other children of the same age would perform.

The second part of the *Register of Progress and Achievement* consisted of lined, blank pages where individual teachers could record, in each subject area:

The results of standardised diagnostic or attainment tests, the results of informal and informal tests, notes or observations on the quality of the pupil's day-to-day performance and comments on a pupil's qualities of stability, co-operation, independence, and perseverance.

Practices in completing this section of the register varied from school to school and within schools, but it was common for results to be represented as grades or marks allowing for easier comparison of students. In line with this positivist approach, assessment was still seen largely as a summative activity. The register instructed teachers to enter the survey results twice a year and other periodic assessments, as described by Black (in Nuttall: 1986), typically comprised a series of *'staccato forms of the old end of session examinations'*.

In addition to the move away from an examination-oriented system of assessment, there are two other key developments indicated in the move away from the Examinations Register and the Proficiency Examination to the Register of Progress and Achievement and the Record Cards. The first relates to the shift away from nationally-driven to teacher-driven assessment. It was Beeby's belief (In Renwick and Ingham, 1974: 157) that teaching is different from most other professional activities in that the knowledge, professionalism and commitment to practice was a valuable resource and lay at the heart of successful curriculum development and delivery. This commitment to the integrity of the professionals included assessment where, in the absence of an external examination, the entire responsibility for monitoring, assessing and evaluating primary student achievement was vested in the professional judgements of teachers.

The third development relates to the encouragement given to teachers to include qualitative as well as quantitative assessment data in assessing students. In the guidelines for completing the Register, teachers are asked to make notes, record observations and write comments. The Record Card Guide states:

Assessments should not wholly come from formal tests. They should comprise all that becomes available to a teacher from day to day performances in the classroom as well as periodical tests.

Despite some indications of a more progressive, learner-focused approach to assessment, the Register of Progress and Achievement and the Primary Record Card were not a serious challenge to the positivist paradigm.

Primary Progress Records and educative assessment

In 1989, a new assessment system was introduced by the then Department of Education. The Register of Progress and Achievement and the Record Cards were replaced by the Primary Progress Records, one for the Junior School (Years 1,2 and 3); one for the Middle School (Years 4,5 and 6) and a final card for the Senior School (Years 7 and 8). They were the product of widespread discussion amongst teachers and the community (ibid) and underwent extensive trialing between 1980 and 1988 in up to 200 schools, prior to their distribution nationwide.

The package that accompanied the cards, *Keeping School Records*, stated a number of reasons for making the change: the inadequacy of the previous system in conveying information about a student's performance and as a guide for future teaching; faulty assumptions underlying the use of the 1 to 5 ratings scale; concern that the ranking system did not lend itself to easy and helpful identification of students with learning difficulties; that labelling students on a ranked scale can be counter-productive; and that there should be a move away from measuring one student's achievement against that of another (Department of Education, 1989b: 17).

The trialing and introduction of the new Primary Progress Records across the 1980s occurred in tandem with major shifts in thinking about the academic achievement, curriculum reform, assessment, the evaluation of teacher competence and the structuring of education (Department of Education, 1987;

Glassman and Glassman, 1988; Lawton, 1989). In assessment in particular, there was what Gipps (1994:1) called '*an explosion of developments*'. These developments gave teachers access to forms of assessment other than norm-referencing, a wider range of assessment tools to use and required that assessment serve a broad range of purposes. Gipps (ibid:10) places many of these developments under the umbrella of an assessment framework that she refers to as educative assessment. There are a number of key tenets that form the basis of paradigm. It focuses on the progress and achievement of individual students and subscribes to the kind of assessment information that can be used by both the teacher and the student to support the learning process. It ensures that assessment is as 'authentic' as possible, in other words, that assessment tasks closely match the desired performance and takes place in an authentic context. It also recognises the importance of metacognition in any assessment process.

The format of the Primary Progress Records and the language of the booklets that accompanied the cards provided considerable evidence of a system that was moving towards educative assessment practices. The booklets stated (Department of Education, 1989b: 5):

1. *The learner should be at the heart of all educational assessment*
2. *Assessment is only useful if it promotes the progress of the learner*
3. *Assessment should be shared with the learner.*

The booklets also referred to identifying a student's strengths, weaknesses and needs and to using the information '*to establish the point they have reached in a particular learning process*' and '*to provide a stepping-off point in developing student-centred programmes*'. There was a clear message that assessment should be ongoing and that information about learners should be collected in

different settings and in a variety of ways. There was a requirement that teachers assess some metacognitive skills and a number of sections that refer to the importance of student self-assessment. The new system placed a much stronger emphasis on formative assessment.

All the summative data on the cards relating to progress and achievement was qualitative. While it was indicated that teachers might collect some quantitative information (eg PAT results and Running Records) for formative purposes and as part of their own assessment system in the classroom, it was not expected that this information be included on the card. Gipps (ibid:15) suggests that educative assessment must resist the tendency to think in simplistic terms about one form of assessment being better than another and that consideration of form without consideration of purpose is wasted effort. It raises the question about whether only qualitative outcomes should have been recorded on the card. Was there valuable data that would have assisted teachers working with students on transition that could not be encapsulated in words? What it did indicate was a complete rejection of the positivist paradigm that had dominated New Zealand classrooms for over a century.

It is also important to note other significant changes that had less to do with assessment and more to do with accountability. The summative statements recorded in each subject area for each student were to relate to specified learning goals against which it was expected students would be assessed. There were goals in knowledge, skills and attitudes for each subject, organised in three levels - junior, middle and senior. This attested to a move away from the complete freedom given to schools and teachers to determine what would be assessed within each curriculum area, which had characterised the era of the Register of Progress and Achievement. As part of the accountability process, schools were also required to carry out twice yearly reviews of the work of classes, syndicates and the school (Department of Education, 1989b: 10). Mention is made of evaluating student needs, standards of achievement, the effectiveness of evaluation procedures and reviewing future programming. Without students' test results and norm-referenced achievement outcomes, the

focus for accountability appeared to be on making the curriculum more transparent and standardised, and on school self-review.

The impact of the New Right, new legislation and the new curriculum

At the end of 1989, a new Education Act was passed that completely altered the structure of New Zealand education. A key plank of the new reforms was self-management, giving the responsibility for the governance and management of individual schools to a parent-elected board of trustees and the principal. This new educational structure entailed a devolution of decision-making in a wide range of administrative areas including resource allocation, staff appointments, support services and staff development. At this point, most decisions relating to the curriculum were still the preserve of the schools.

Two years after the introduction of the structural reforms, the Ministry began a major overhaul of the national curriculum, the biggest rejuvenation since the 1940's. This culminated in the release of the National Curriculum Framework in 1993 and the subsequent gazetting of new curricula, on a drip-feed basis. In 1998, this process is almost complete. Thus, the national curriculum is centrally determined and although the board, the principal and staff have considerable control over the way in which it is implemented, there is a high level of curriculum accountability built into the brief of the Education Review Office (ERO)⁵.

The extent and direction of the reforms, both here and in countries such as the United Kingdom and the USA, have been influenced by the ideologies of the New Right which have linked educational performance to economic growth. As Goodson (1990:230) has argued in relation to the United Kingdom:

⁵ The Education Review Office (ERO) is an independent body that reports directly to the Minister of Education and to Parliament. It is responsible for reviewing the performance of early childhood centres and schools on a regular basis. It produces a written report at the conclusion of each review and these reports are made public.

The introduction of the national curriculum in the UK has been linked to the problems of national economic decline and the belief that curriculum coordination will aid a project of national regeneration.

In New Zealand, a similar view was articulated by the Minister of Education. In a speech to a teachers' conference in 1991 he stated that:

The change [to increased Government involvement in assessment and curriculum matters] is a result of the Government's heightened recognition of education as a significant aspect of national development, its central position in the development of a sound economic strategy.

This was reiterated by the Ministry when expressing its broad objectives in the *National Curriculum* (1991: 1,19). It notes that the purpose of the new curriculum was to:

...enable students to take their full place in society and to succeed in the modern competitive economy...[and to] participate effectively and productively as responsible and informed citizens of New Zealand's democratic society and economy.

Within the context of New Right thought, the key to improving education is by creating an education market in which parental choice and competition trigger the drive to improved standards. In this climate, assessment plays a pivotal role because it can be used as a 'market signal' (Willis, 1992: 205) - a means whereby the educational performance of schools and teachers, as well as students, can be judged.

Willis (ibid) argues that, while there is evidence that the New Right has strongly influenced the administrative restructuring of the nature of accountability, it has had less influence on issues relating to pedagogy, the curriculum and assessment. In relation to assessment, she cites as evidence the outcomes of

the Able Project⁶ and the document they produced, *Tomorrow's Standards* (1990). This document clearly rejected references to models for accountability in favour of principles to support better learning. For example:

The primary purpose of assessment should be to provide information that can be used to identify strengths and guide improvement. (Principle No.2)

Emphasis should be given to identifying and reporting educational growth and progress, rather than to comparisons of individuals and schools. (Principle No. 3) (ibid:8)

In the early 1990's, teachers also continued to use the Primary Progress Records and to reflect the student-centred philosophies and individualised approach promoted in *Keeping School Records*. Even the New Zealand Curriculum Framework (1993: 24) stated:

The primary purpose of school-based assessment is to improve learning and the quality of the learning programmes.

Later, in 1995, in response to concerns expressed by principals and teachers that assessment practices were becoming mechanistic and that the curriculum was becoming assessment-driven, the Ministry was persuaded to publish a statement in the *Education Gazette* (May, 1995) that clearly articulated the importance of using assessment to support and enhance student learning. The article declared that:

The Ministry of Education is concerned that assessment does not dominate the teaching and learning process, rather, it should be an integral part of teaching and learning. ...Assessment should provide

⁶ Able is an acronym for Assessment for Better Learning. The main purpose of the Able Project was to provide the Government with models and procedures to monitor the effectiveness of the New Zealand school system on students' learning and to assess the effect of individual schools on students' learning achievements.

teachers, parents and students with useful information about students progress. ...Schools should aim to develop assessment procedures which are manageable for teachers, non-intrusive for students, and focussed on promoting learning. ...Teachers should be encouraged to value their professional judgement in assessing student achievement as it is this judgement which provides the most important basis for effective school assessment.

However, the rhetoric appears to contradict the legal requirements for assessment, introduced by the Government in 1993, in the form of the National Administration Guidelines (Ministry of Education, 1993b; see Appendix A). They indicate a move towards the recapturing of assessment for the purposes of monitoring and accountability:

Each Board, through the principal and staff, will be required to:

- i. implement learning programmes based upon the underlying principles, stated in the essential learning areas and skills, **the national achievement objectives**;*
- ii. **monitor progress** against the national achievement objectives;*
- iii. assess student **achievement**, maintain individual **records** and **report** on student progress (my emphasis).*

The language used in the Guidelines - 'national achievement objectives', 'monitoring progress', 'achievement', 'records' and 'reporting' - contrasts sharply with the references to the purposes of assessment in the other Ministerial documentation mentioned earlier (Keeping School Records and The Curriculum Framework). It insinuates that learning and the quality of learning programmes will be improved through accountability rather than through integrating assessment with learning.

This interpretation is supported by the Education Review Office, the body responsible for ensuring that schools are meeting the requirement of the NAGS. They outline the purpose of assessment as (Education Review Office, 1995: 9):

...to provide information to students and parents; to inform teaching and learning programmes; to inform policy and decision makers such as Government Ministers and boards of trustees.

As Hill (1997: 3) points out, government regulation and the main agency responsible for checking that regulations are implemented, are framing assessment more as a vital element of systemic control than as a way to improve the quality of learning.

This analysis suggests Willis is perhaps mistaken in proposing that assessment has not been captured by the New Right. What emerges is a tension between the requirement of accountability and the desire that assessment should promote learning. Perhaps it should not be surprising that Wylie (1997) reported that more schools (66%) rated change to assessment practices as the most significant change they have had to come to terms with since the introduction of 'Tomorrow's Schools'⁷. Other related issues, such as the monitoring and evaluation of school and class programmes and reporting student achievement to parents were also noted among the seven major changes reported (ibid: viii).

Summary

This chapter has provided an historical context for the curriculum and the assessment initiatives introduced into New Zealand primary schools in 1993. It has highlighted the emphasis previously placed on an examination system, and later on norm-referencing, as a way of measuring student progress and

⁷ Tomorrow's Schools was the name given to the educational reforms introduced in 1989.

achievement. What has also been brought into focus, is the high degree of professional autonomy given to New Zealand teachers and schools to make judgements about the progress and achievement of individual students and about how that data might inform programme evaluation and school-wide practices. It outlined the tensions inherent in assessing for purposes of supporting student learning and for accountability purposes. The next chapter places the discussion and debate in the context of both the New Zealand and international literature on assessment.

CHAPTER THREE

LITERATURE REVIEW

'If assessment has an educational role to play it is a good deal more complex than even well-informed educational opinion would seem to believe, and we do ourselves a disservice to pretend otherwise' (Torrance, 1993: 333).

Introduction

This thesis involves the integration of substantive and process issues in describing the development of new assessment systems within an action research methodology. The methodological processes are outlined in the next two chapters. This chapter reviews the literature in terms of the substantive issues that relate to the involvement of the four schools in the Assessment for Better Learning (ABLE) Contract.

As an area of study within education, assessment (as opposed to educational measurement dominated by psychometric testing) has a relatively short history. Major developments have occurred in the field since the beginning of the 1980's and this burgeoning interest has resulted in a wealth of literature on the topic (Earl and LeMahieu, 1997: 150; Gipps, 1994: 1). Most of the literature has emanated from Britain and the United States and, in comparison, there is not a great deal that is New Zealand based. While the previous chapter provided an historical context of assessment practice in New Zealand, this chapter will put the study into the context of international research and development.

Firstly, the chapter describes the interaction between assessment and the development of the different theories of how students learn. It explores the moves away from norm-referencing towards standards-based and performance assessments and to creating authentic tasks and contexts for assessment. It

details an expanded repertoire of how assessment outcomes can be used and describes the tension between the often competing philosophical approaches of formative and summative assessment. Towards the end of the chapter, the shift from a teacher-dominated process towards greater student and parent involvement is discussed and, finally, the implications of new reporting requirements. These issues are explored under five broad headings:

1. The interaction of assessment with theories of learning
2. The role of assessment
3. Approaches to assessment
4. Designing assessment tasks and issues of validity
5. The role of teachers in the assessment process
6. Reporting to parents, trustees and government agencies

The interaction of assessment with theories of learning

As mentioned in the previous chapter, assessment has undergone some major paradigm shifts since testing and measurement was first introduced into education over a century ago. Gipps (1994) describes three assessment paradigms:

- **The psychometric paradigm**, which is largely concerned with formulae, quantification, accuracy and objectivity
- **The educational measurement paradigm** which is about asking 'how well' rather than 'how many' and aims to devise tests which look at the individual as an individual rather than in relation to other individuals and to use the information to aid educational progress
- **The educational assessment paradigm** which also subscribes to the view that assessment must support learning but which aspires to the use of a broader range of assessment strategies to assess a broader body of cognitive aspects than just mere subject-matter acquisition and

retention. It embraces the view that assessment should be authentic, in other words, where the assessment closely matches the desired performance and takes place in an authentic, or classroom, context.

The impetus for rethinking assessment practices and shaping an educational assessment paradigm is closely linked to a reconceptualised model of learning. The traditional model, on which psychometric testing is based, assumes that knowledge and skills can be compartmentalised and that complex competencies can be broken down and learned separately. Called behaviourism, learning is believed to be linear and sequential and that complex understandings only occur when elemental prerequisite learnings are mastered (Resnick and Resnick, 1992: 42). Another assumption is that once a skill is learned, it can readily be transferred from that context to another (ibid, 43).

An alternative learning model comes from recent work in cognitive and constructivist psychology. Gipps (1994: 21) describes this model as:

show[ing] learning in terms of networks with connections in many directions; not of an external map that is transposed directly into the student's head, but an organic process of reorganising and restructuring as the student learns.

In other words, the learner is actively engaged in making sense of any new knowledge. Knowledge acquisition is seen as something cohesive and holistic which provides a 'scaffolding' for later learning. Both models hold a very different view of the learner. In the behaviourist's mind, the learner is a passive absorber of information and facts provided by the teacher. Learning is incidental and school work is a requirement not necessarily related to knowledge and learning (ibid, 23). For the constructivist, both the teacher and the learner are engaged in knowledge construction and metacognition - learning how to learn - is a critical component of the process. In this frame, learning is an intentional and contextualised process.

This more expansive view of how students learn meant reshaping the way both academics and practitioners thought about assessment. No longer could assessment atomise and compartmentalise knowledge. It was no longer satisfactory to just test recall of facts. Assessment needed to find ways of ascertaining the level and complexity of a student's understanding and, most importantly, to do this in ways that promoted further learning (Piper, 1997: 95).

Alongside the behaviourist/constructivist debate, other developments in cognitive science were also impacting on the way models of assessment were being conceptualised. The work of Vygotsky (Bruner, 1985) on the *zone of proximal development* (ZPD) identified a gap between the actual development level shown by the child's unaided performance and his or her potential under adult guidance or in collaboration with more capable peers. Called scaffolding, this concept was built on by Feuerstein and others in a range of assessment approaches called guided or dynamic assessments (Brown et al, 1992: 139). All are based on close observation, diagnosis and teacher/student interaction with a clear focus of tapping into the learning potential of a student (Gipps, 1994: 27). While much of this work is based on research with individuals rather than whole classrooms, the concepts have been integrated into assessment theory and practice. Performance-based assessment tasks, collaborative assessments and even allowing students access to resources in a testing situation are all examples of their influence.

The Resnicks (1987, 1992) described the challenge of 'the thinking curriculum', a term that refers to the 'new basics' of thinking, reasoning and learning how to learn. Although these new basics have become important to global changes in communication, technology and in economic terms, they also have resulted from work in higher order thinking skills: metacognition (Gipps, 1994: 25). An essential aspect of metacognition is that learners help to plan, monitor, orchestrate and control their own learning. In this model of learning, assessment is not a matter of using one-off, multiple choice or short answer type tests to check on whether information has been received. Understanding also involves how knowledge interlinks, its applicability, invention and critique.

No one procedure will measure all these characteristics well (White, 1992: 161) and what teachers need is a diverse repertoire of assessment techniques to draw upon.

In the same way that notions of **what** should be assessed have expanded, so the **purposes** and **uses** made of assessment have also diversified. The next section explores this diversification and describes the inherent tensions between the different purposes.

The role of assessment

Assessment in education has been described by Harlen et al (1992) as the process of gathering, interpreting, recording and using information about students' responses to an educational task. At one end of the continuum the assessment task may be normal classroom work and the process of gathering information or listening to what the student has to say. At the other end of the continuum, the task may be a written, timed examination which is read and marked according to certain rules and regulations. In the same way that the task and the setting may be different, it also fulfils a number of readily identifiable and different functions (Rowntree, 1989: 15; Harlen et al, 1992: 217; Black, 1993: 50; Eisner, 1993:224):

1. Identifying the particular achievement levels of individual students so that decisions can be made about which class and/or which particular group they will work within (placement and, sometimes, gatekeeping function).
2. Identifying the particular or learning needs of individual students and groups of students (diagnostic function).
3. Providing feedback to teachers and students about ongoing progress in learning (formative function).

4. Communicating the nature and level of students' achievements at various points in the year or in their schooling - to parents and on transition to another class or school (summative function).
5. Providing feedback on the quality and effectiveness of the programme being provided (evaluative function).
6. Summarising what has been achieved for the purposes of selection and qualification (certification role).
7. Judging the effectiveness of educational institutions and the teachers within them and the effectiveness of the system as a whole (accountability and quality control role).

Eisner argues (ibid, 225) that the foregoing functions can be reduced to three major areas of focus - programme evaluation, teacher evaluation and student evaluation - and that different forms of assessment and evaluation are required for the different functions. The implication is a growing pluralism in method and knowledge that will dramatically increase the array of data describing educational practice and its consequences which Eisner believes will complicate rather than simplify our understanding of schooling.

For simple conclusions, one wants simple data or data arrayed on a common metric. When neither the data are simple nor the metric common, complexity is virtually inevitable.

Black (1993), Eisner (1993), Crooks (1988) and Madaus (1988), all argue that most of the investment in assessment and testing, in research and development and in classroom practice, has been devoted to its summative, accountability and certification functions. In contrast, the formative function, where teachers and their students use their own and each others' assessments to adapt programmes to better meet the needs of the students, has largely

been ignored. Yet Crooks (ibid) points out, that students spend vastly greater amounts of time engaged in classroom evaluation activities than in standardised testing. Black and Wiliam (1998) also provide compelling evidence that the educational and psychological effects of formative classroom assessment are generally substantially greater than the corresponding effects of paper-and-pencil tests that offer reliability and easier aggregation.

So, what are the compelling features of formative assessment that make it crucial to effective teaching and learning practice? Sadler (1989: 123) places emphasis on students having a clear notion of the standards or goals to be reached, being able to compare their actual performance with the desired performance and engaging in the appropriate action to close the gap. Crooks (ibid) also states that research has repeatedly demonstrated that students achieve most and gain most on key motivational variables when evaluation standards are high but attainable. In addition, he places weight on the frequency of assessment whereby students are given regular opportunities to practice the skills and knowledge that are the goals of the programme and to obtain feedback on their performance.

Black and Wiliam (1998), Tunstall and Gipps (1995), Gipps (1994) and Crooks (ibid) all highlight the importance of giving useful feedback to students. They suggest that teacher feedback not only aids the learning process but also affects how students view themselves: students' academic self-esteem is determined largely through feedback from teachers. Feedback needs to focus students' attention on their progress in mastering educational tasks, it should be given when it is still clearly relevant and should be specific and detailed enough to be of use to the student.

There is much controversy about the relationship between formative and summative functions (Black and Wiliam, 1998: 17). One position is that the two are incompatible and that assessment instruments and procedures cannot serve two functions without weakening the efficacy of both (Murphy, 1990; Harlen et al (1992). Black (1993: 61) raises two concerns about this view.

Firstly, the writers do not explain how the need for summative assessment will be satisfied, for example, whether it should be external or school-based. Secondly, no account is taken of the propensity for the summative process to swamp the formative work. Resnick and Resnick (1992:57) support this view with their contention that:

Whether we like it or not, what is taught and what is tested are intimately related. No serious possibility exists for creating accountability tests that will not eventually influence what is taught and how it is taught in the schools. This means that there is no way to create accountability tests that will be curriculum-neutral.

The alternative view to Black (ibid) and Resnick and Resnick (ibid) posits that while formative assessment must be pursued for its main purpose of providing feedback into the learning process, selection and aggregation procedures applied to the formative process can also be used to produce information for summative purposes. While this would protect the assessment process from being swamped by 'high stakes' testing, Black (ibid) worries about the conflicting advisory and adjudicating roles of the teacher and the whether detailed and fine-grained formative data can be extrapolated to provide a final outcome that is fair to the student. To counter this, Harlen et al (1992: 223) make a useful distinction here between the 'summing up' and 'checking up' function of summative assessment. In the former, information collected over a period is simply 'summed' every so often to see how students are getting on. Checking up, however, is when summative assessment is done through the use of tests or tasks specifically designed for the purpose of recording performance at a particular time. The writers maintain that formative assessment can be used for summing up purposes without compromising its feedback role. This view might meet the needs of the policy makers in this country who are demanding access to standardised data that would enable them to compare the quality of individual schools and to ascertain the overall health of the nation's education system.

Torrance (1993: 335) warns of the dangers of 'dumbing down' formative assessment. He reminds readers that formative assessment could easily be construed as an essentially behaviourist activity in the tradition of mastery learning - define your objectives and teach them quite specifically, making sure that teachers and students alike know what behaviour is required of them, ie what counts as achieving the objective. Torrance (ibid) argues that for formative assessment to fulfil its promise, it must sit within a constructivist perspective. Learning should be 'scaffolded', by students being set appropriate tasks and being provided with appropriate support, with the purpose and focus of assessment being to identify what it is that the students could achieve next.

In the same way that assessment can serve different purposes, there are also distinct approaches to assessment that shape how tasks are constructed, that determine the context in which they are used, how the data are portrayed and how the outcomes are used. The discussion now turns to the impact of these differing approaches and their links with the theories of learning that have developed over time.

Approaches to assessment

The traditional approach was that of norm-referencing which was discussed earlier in Chapter 2. In summary, it was based on technicist assumptions and was designed to enable comparative judgements to be made, student against student, student against the norm. It was not designed to generate specific information about what an individual child knows, understands and can do, irrespective of other children (Sutton, 1991: 4). For these reasons, a national, norm-referenced assessment system disappeared from New Zealand primary schools in the 1980's.

Linked with the advent of behaviourism, criterion-referenced assessment emerged on the education scene in the 1960's (Kulik et al, 1990: 265). Its rise in prominence is linked with the publication of Glaser's seminal paper on criterion-referenced testing in 1963 (Gipps, 1994:79). It was also the basis for

the development of mastery learning of which there are a number of variations (Ritchie and Carr, 1992: 191). Criterion-referenced assessment differs from norm-referenced assessment in that a student's performance is measured against pre-determined expectations. These are usually written down and built into the assessment process. The assessment procedure is then designed to provide evidence for each child, of whether those specific expectations, or criteria for success, have been met (Sutton, *ibid*).

A difficulty with criterion-referenced assessment is the degree of specificity required. If the expectations are highly specific, which may make the task easier to judge, it is likely there will be more of them and that teachers will use them as their main learning targets. On the other hand, it is more difficult to dependably assess expectations that are expressed in more holistic ways: the more holistic the learning, the more interdependent the different parts of learning and, therefore, the more difficult it is to break down the task into specific and discrete items (Elley, 1995: 79). Even if that occurs, what is being measured are the fragmented bits, not how the learner manages to put them together. Despite these issues, Kulik et al (*ibid*: 292), in their meta-analysis of the effectiveness of 108 controlled evaluations of mastery programmes, give clear evidence that impressive achievement gains can be made with such programmes, especially by weaker students.

Gipps (1994: 85) discusses the difficulties in aggregating criterion-referenced information, where a detailed performance profile may have to be 'collapsed' into a figure or a grade for reporting purposes. She advocates the use of more descriptive summaries and refers to the work of Hambleton and Rogers (1991, in Gipps, *ibid*) in utilising expertise in setting cut-off scores. Elley (*ibid*: 93) cites a number of variables in criterion-referenced test design and administration (eg generalising from a specific test question to the particular criteria or standard that they purport to test; maintaining equivalence between different testing 'sites') that he believes make a strong case for not aggregating data for purposes beyond the classroom or the school.

A derivation of criterion-referenced assessment is standards-based assessment. Conceptualised in the main by Sadler (1992) it differs from criterion-referenced assessment in that it moves away from tightly specified criteria and does not rely on sophisticated statistical and technological expertise for grading and assessing. Standards-based assessment draws upon the professional ability of competent teachers to make qualitative judgements, which they do in every-day teaching, and for a pattern of performances over a series of testing episodes and tasks to take precedence over a final score or grade (Gipps, *ibid*). Specification of the standards comes from verbal descriptions and exemplars (Diez and Moon, 1992: 40). Sadler (*ibid*: 4) argues that this approach is consistent with the educational assessment model in that the locus of control remains with the teacher and the student.

Performance assessment calls for students to demonstrate their capabilities directly, by creating some product or engaging in some activity (Haertal, 1991: 3). It aims to model real learning activities and, as such, occurs in highly contextualised situations (Wiggins, 1993: 208). Evaluating student achievement through performance assessments is not a new strategy (Khatttri et al, 1995: 80). Good teachers have always judged and monitored their students' progress through observations, experiments, written assignments and research reports. What is new, is the use of performance assessments for more formal testing situations and a systematic shift away from paper-and-pencil tests for measuring instruction and for accountability.

Gipps (1994: 99) outlines a number of characteristics that apply to this approach to assessment: it is time-consuming; tends to provide detailed multi-dimensional information about a particular skill or area; because of the time factor, depth may be exchanged for breadth; scoring is generally complex and usually involves the classroom teacher; standardisation of the performance is difficult and therefore reliability, in the traditional sense, is not high.

Nonetheless, there is a growing body of evidence that demonstrates that performance assessment provides a real opportunity to improve teaching and learning (Wolf et al, 1991; Gardner, 1992; Wiggins, 1993; Khattri et al, 1995) and, despite the difficulties in design and scoring, can provide reliable measures that can be used beyond the classroom for accountability purposes (Moss, 1992; Resnick and Resnick: 1992). Research by Khattri et al (ibid) also shows that, while performance assessments have had a marginal effect on the curriculum teachers use in their classrooms, mainly because the curriculum was prescribed, the impact on instruction and teacher roles in that process has been considerable.

Performance and authentic assessment are terms that are often used interchangeably. Although both labels might appropriately apply to a similar approach to assessment, they are not synonymous. Meyer (1992: 39) refers to performance assessment as an indicator of the kind of student response to be examined; authentic assessment refers to the context in which that response is performed. While not all performance assessments are authentic, it is difficult to imagine an authentic assessment that would not also be a performance assessment. Nonetheless, for assessment to be classified as authentic, the task must be carried out in an authentic context and be a part of the normal classroom programme. The most typical example of authentic assessment is the portfolio which contains examples of actual student performances often presented chronologically and elicited under normal classroom conditions in the classroom context.

The final approach to be discussed is self-assessment. This is described by Towler and Broadfoot (1992: 137) as the process of reviewing past experience, seeking to remember and understand what took place and attempting to gain a clear idea of what has to be learned and achieved. They argue that self-assessment helps children develop the meta-cognitive skills of reflection and self-criticism and encourages motivation by giving responsibility to children for their learning and by signalling that their opinions are important. While it could be argued that self-assessment is a learning tool not an assessment tool,

Towler and Broadfoot (ibid) point to research that suggests otherwise. They propose that self-assessment can also be used to circumvent time-consuming record-keeping which is often mandated for teachers and which nobody uses. In their view, Records of Achievement based on on-going self-assessment and negotiation, beginning at 5 years old, could more adequately meet the need of providing quality information on transition from class to class each year and between phases of schooling.

While the approaches to assessment provide a framework for assessment practice, the next section addresses more practical issues in relation to the design of assessment tasks. It then relates these to the implications for classroom practitioners and, by inference, to school administrators.

Designing assessment tasks and questions of validity

Textbooks and articles abound with the principles that should underpin good assessment design. Given that modern assessment practice predominantly subscribes to the paradigm of educational assessment, it would be useful, at this point, to summarise some key criteria. Using, what Eisner (1993: 226) describes as eight essential criteria as a basis, the following points synthesise eleven key principles critical to both creating and appraising assessment tasks. The tasks should:

- assess what really matters
- assess what students know and can do
- reflect the tasks they will encounter in the world outside the school
- reveal how students go about solving a problem, not only the solutions they formulate

- help students demonstrate that they have connected their ideas to a larger intellectual context and that their knowledge is not just a collection of fragmented bits and pieces of information
- make it possible to display intelligent application and adaptation of the ideas learned (assessment tasks that require recapitulating content are, at best, only a test of memory)
- permit the student to choose a form of representation to display what has been learned.
- make possible more than one acceptable solution to a problem and more than one acceptable answer to a question
- not necessarily limit assessment to a single performance
- encourage the development of self-assessment skills
- provide a place for assessing progress on group tasks and the realisation of group goals

In addition to these principles, considerations pertaining to validity are also crucial to assessment task design. Crooks et al (1996: 1) state that, in the past, much more emphasis has been placed on reliability and generalisability than on validity. They maintain that this has largely been attributable to the algorithmic nature of procedures for reliability and generalisability which are much easier to work with than estimation of validity. The latter relies more on human judgement and this makes it harder to test for, report on and defend. In addition, the complexity of what is now expected to be measured further complicates the process. Crooks et al (ibid) worry that the complexities associated with validity could threaten the investigation of new approaches and the resolution of issues that surround its use.

In a major review, Moss (1992) identifies a number of sets of criteria that can provide guidance in establishing validity (Messick, 1989; Cronbach, 1988; Fredriksen and Collins, 1989; Haertel, 1990; Linn, Baker and Dunbar, 1991). Crooks et al (ibid) also posit a set of criteria that identifies a number of threats that can occur at any point in the assessment procedure - in administering the test, scoring, aggregating, generalising, evaluating, making decisions and assessing the impact on the student. They advocate that much of the evidence for validation should be developed at the planning and design stages and be later verified and supplemented after putting the test into practice.

Even though Crooks et al had larger scale testing in mind when describing their validation criteria, there is little evidence to suggest that any of these issues are at the forefront of the minds of practitioners and school administrators when they are preparing assessment tasks for a class, a cohort or across the school. When examining the impact of classroom evaluations on students, Crooks (1989: 440) cites research that suggests that, while teachers judge evaluative activities to be important aspects of teaching and learning, they are often concerned about the perceived inadequacies of their efforts. He concluded that teachers have little or no formal training in educational measurement techniques and many who do, find it of little relevance to their classroom practice. His review (ibid) also showed a need for improved teacher skills in observation and non-test means of evaluation.

Crooks (ibid) provides considerable evidence that teacher-made tests often fail to reflect teachers' stated instructional objectives and tend to emphasise lower order cognitive skills. Redding (1992: 49) supports this when she describes the difficulties teachers have in assessing higher order learning objectives such as problem solving and being collaborative and self-directing:

Thinking up the assessment tasks was easy, but specifying the criteria and quality standards by which student performance could be judged was beyond our expertise.

Gipps (1994: 160) goes further and states that:

Teachers cannot assess well subject matter that they do not understand, just as they cannot teach it well. Teachers have to understand the constructs which they are assessing (and therefore know what tasks to set); they have to know how to get at the pupil's knowledge and understanding (and therefore what sort of questions to ask); and how to elicit the pupil's best performance (which depends on the physical, social and intellectual context in which the assessment takes place).

Willis (1994: 166) supports Gipps when she concludes that the work of teachers lies at the heart of assessment reform. Her rationale is that teachers, with their students, are responsible for making the links between curriculum, learning and assessment. Therefore, they need to fully understand all aspects of the assessment process. This includes an ability to use a range of assessment techniques that can be appropriately, and validly, matched to what is being assessed, combined with skills in marking and moderation. They also need to have an appreciation of the different approaches to assessment and an understanding of the ideological and theoretical underpinnings of the different assessment models (Blackmore, 1988: 49). The critical function of the teacher in the assessment process will now be discussed in greater detail.

The role of teachers in the assessment process

When investigating what occurred in classrooms when a compulsory assessment system was introduced into English schools, Gipps et al (1995: 20) reported that teachers found it very difficult to articulate how they would do, or had done, their teacher assessments. They found that:

Many of the interviews yielded vague descriptions of collecting evidence and details of record-keeping and planning (of assessments). Teachers found it difficult to describe precisely what they used to determine levels of attainment and how they reached this decision.

Hill (1997:16) describes wide variations in the ways teachers carry out assessment in their classrooms and the roles they play in the assessment process. She posits that the reasons for these variations could be linked to one or a combination of factors to do with the experience of the teacher, their familiarity with the curriculum, their personal views about how learning occurs and possibly even the age group of the children they teach. Evidence suggests (Gipps et al, 1995: 36; Hill, *ibid*) that there are, broadly speaking, three types of assessors:

1. integrated, systematic assessors or trackers;
2. unit assessors, evidence gatherers or checkers
3. head-note assessors, intuitives or markers

Typically, the integrated and systematic assessors are able to plan for individual differentiations within their classes. They have the ability to cover a full range of specific learning outcomes which they are able to integrate with specified national curriculum objectives. Their data gathering, decision making and recording practices are ongoing, diagnostic and formative.

The unit assessors, evidence gatherers and checkers have achieving unit objectives as their priority. They also cover specific learning outcomes relevant to their students and national curriculum objectives but each tends to be treated separately rather than integrated into the teaching and learning process. Their practices are more summative in nature and outcomes are recorded at the end of a unit or at the end of the term. There appears to be less assessment-related, teacher self-reflection and analysis, and less teacher/student interaction on assessment eg use of questions and giving students feedback on their work.

The head note assessors, intuitives and markers teach to the national curriculum objectives and record achievement periodically. Like the unit assessors, their assessment is largely summative but less frequent, maybe at the end of each term or twice yearly. Their decisions are more likely than with

either of the other two types, to be based on head notes. While it could be argued that the professional judgements of integrated, systematic assessors are no different from the head notes of the third group, the difference seems to lie in the depth and integration of analysis and evidence that is brought to bear on the judgments made by the first group.

Quite clearly, different teachers will play different roles in their classrooms, both in the way they teach and the way they assess. They will also bring the beliefs and assumptions that apply to those roles, to any professional and school development initiatives.

In the view of Darling-Hammond (1994: 497), the involvement of teachers in the development and implementation of new assessment practices is just as important as the nature of the assessment tools and strategies. If assessment practices are to improve, teachers need to engage in a dynamic process of staff and school development (Willis and Bourke, 1998). Assessments that are externally developed and scored will not transform the knowledge and understanding of teachers. Instead, teachers must actively engage in collaboratively defining, redefining, testing, and activating their own constructed and contextualised understanding of what is worth knowing and how it is worth assessing.

The findings suggest (Khatti et al, 1995: 81) that as teachers learn more about what students know and think, as well as about how they learn, the more likely it is to affect their pedagogy. Education reformers see the role of teachers changing from being a disseminator of information to a facilitator of learning and Katti et al (ibid) believe that the use of assessments that match these new models of learning, such as performance assessments, are assisting teachers to make this kind of adjustment.

Crooks (1989: 467) is less optimistic and maintains that insufficient credence is given to the impact of classroom evaluation on students. He suggests a much higher level of collaboration between teachers and schools is needed.

A more professional approach to evaluation would demand regular and thoughtful analysis by teachers of their personal evaluation practices, greater use of peer review procedures and considerable attention to the establishment of more consistent progressions of expectations and criteria among educational institutions.

In the same way that Crooks is advocating greater collaboration between teachers, schools and other educational institutions, there is an increasing body of research suggesting that changes are needed in the ways schools report to parents on the progress and achievement levels of their children.

Reporting to parents

Guskey (1995) writes that grading and reporting on student learning remains a favourite topic for researchers. To illustrate his point, he describes a recent review of the ERIC system⁸ found over 4000 references to journal articles and reports on the topic that had been published since 1960. Guskey (ibid) argues that teachers do not need grades and report forms to teach well and students can, and do, learn without them. He suggests that the primary purpose of grading and reporting is other than the facilitation of teaching and learning. This represents a traditional, and somewhat narrow view of reporting, given the large body of evidence (Ramsey et al, 1993; Earl and Le Mahieu, 1997: 156) that demonstrates the benefits, for all of the stakeholders, of teachers and parents working together to support children's learning. Reporting processes must surely offer an opportunity for this to occur.

⁸ ERIC is an acronym for Educational Resources Information Centre, a library reference resource for educators.

This view is supported by recent Phi Delta Kappan/Gallup Polls (Elam, Rose and Gallup, 1991; 1992; 1994) which provide ample evidence that, not only do parents and large percentages of the community care about how well students and schools are performing, but they also want to know more. The Polls indicate that they want better information about how their children are performing, more informative formats for reporting on progress and they want to be able to use this information to understand their child's strengths and weaknesses.

A variety of reporting methods are now available to schools. They include letter grades, numerical scores, developmental or proficiency scores, checklists, written narrative reports, portfolios and verbal interviews or conferences with or without students (Bailey and McTigue, 1995: 121). Bailey and McTigue (ibid: 119) remind us that any communication system consists of four components: *purpose* (why it is being communicated); *content* (what is being communicated); *process* (how it is being communicated); and *audience* (to whom it is being communicated). On this basis, Wiggins (ibid, 141) suggests that report cards, or any other reporting mechanisms, are much more than a design and graphics problem. He believes that many well-intentioned schools keep reinventing reporting systems based more on their own interests than those of their various clients.

Guskey (ibid) laments that often schools attempt to address all the variations within the four communication system components (purpose, content, process and audience) with a single method and end up achieving none very well. For example, letter or numerical grades are a relatively quick and easy process for teachers to use but require such abstraction of information into a single symbol that it may not adequately reflect a student's progress or achievement. Conversely, narratives and checklists often do not communicate whether a student's progress measures with that expected of the level. Narratives are also time consuming for teachers to prepare.

Wiggins (1995: 148) alerts us to a number of anomalies that can occur when assigning grades. He draws a distinction between grades assigned on the basis of expectations of students, given the larger context of their peer group, and those assigned on the basis of each student's personal circumstances. The difference is that the former is a 'normed' grade and the latter is a growth grade. He warns that both should be reported on separately as they represent distinctly different judgements. Concern is also expressed at the tendency for teachers to confuse effort and other attitudinal traits with achievement (Wiggins, *ibid*; Stiggins, 1994: 363). Both authors note that distinctions between effort, progress and achievement are rarely made, even though the differences are critical when considering a student's record of learning.

Brookhart (1993: 139) is concerned that the assigning of grades should not simply become a sterile process of following measurement instructions. He argues that teacher subjectivity is beneficial rather than detrimental to the process. Because teachers know their students, understand the various dimensions of their work and have clear notions of the progress made, their subjective perceptions yield very accurate descriptions of what students have learned. This can also pose a dilemma for some teachers. Because grading and reporting is, typically, a summative exercise, it involves judging the adequacy of a student's performance at a particular point in time. Brookhart (*ibid*: 140) cites a number of studies that indicate that if a teacher's first priority is to be an advocate for the student, concerns about the impact of the grading process on the student can influence how grades are assigned. In other words, the dual roles of being both advocate and judge, are not necessarily compatible for teachers. Bishop (1992: 16) goes as far as to say that this is a rationale for the external testing of students.

Traditional parent-teacher interviews are a common way of reporting students' progress. There is some documented evidence that teachers often see these interviews and the process of preparing reports, as adding a lot of extra time and paper work to an already demanding teaching job (Little and Allan, 1989, 210). There is a growing incidence of some schools involving students in the

interview or conference process (Le Countryman and Schroeder, 1996: 64; Bailey and McTigue, 1995: 136), often using portfolios (Bailey and McTigue, *ibid*). Little and Allan (*ibid*) claim that student-led conferences are an opportunity for students to take direct responsibility for their own learning as well as for reporting on their own progress. However, difficulties cited with this process include preparation time for students and teachers, finding time in the schedule and the actual amount of conference time needed.

Brewer and Kallick (1996: 178) and Renwick and Gray (1995:49) discuss the possible impact of technology on reporting and recording practices in the future. While they were able to cite some examples of schools using technology to support their reporting systems eg digital portfolios, video or computer disk report cards, the scenarios they describe, for the most part, are visionary. They do not restrict technology to the computer but, rather, envisage a multi-media impact on the whole of assessment not just reporting. For a long time, student achievement was only observed by those who happened to be in the classroom at the time. Technology has the potential to allow student achievement to be more visible and for data to be retrievable at any point in time. Combined with higher demands for customer service and accountability, Brewer and Kallick (*ibid*: 181) predict that it is inevitable that schools will adopt technology to support their reporting practices. Nonetheless, lack of parent and teacher skills, the cost of hardware and access to easy-to-use software are just some of the major barriers to be overcome in making this a reality for New Zealand schools.

Summary

There is a clear link between the development of theories of learning and major shifts in the models of assessment practices that have dominated the curriculum and the practice of teachers in classrooms. This review has highlighted the benefits to students of assessment practices that are based on an educative paradigm with a strong focus on formative assessment, which is based on sound criteria and places an emphasis on performance assessment

and the assessment of metacognitive skills. There is little evidence of the extent to which New Zealand schools are operating within the constructs and practices of an educative paradigm. This will be an important focus of this particular research exercise.

Some insights have been gained into the changing role of teachers in assessment and into a number of issues critical to understanding the kind of professional development that is needed if assessment is to continue to support and enhance student achievement and, more importantly, assist them in learning how to learn. Questions remain about the role that New Zealand teachers believe they have in the assessment process and of the particular knowledge they have, or do not have, that supports or hinders assessment of student learning. Very little is also documented about how teachers, including those in New Zealand, use other teachers' assessment information, particularly when students transfer from one class to another or from one school to another.

There is also little research into how New Zealand teachers are applying what they already know about assessment, to a new national curriculum and new legislative requirements. In Chapter 6, the initiatives, struggles and insights of the four schools in the study are outlined as they come to terms with these a range of mandated requirements. Because the issues related to these developments were relatively unexplored, an action research methodology seemed ideally suited to investigating what these issues might be and to supporting the schools in seeking resolutions. A detailed description of the theory and practice of action research methodology is the subject of the next two chapters.

CHAPTER FOUR

THE METHODOLOGY IN THEORY

'Research that produces nothing but books will not suffice' (Lewin, 1948).

Introduction

The prime purpose of this chapter is to provide a methodological framework against which an action research process can be measured in practice. It describes the philosophy and key principles that underpin action research and critiques the models and tools that can be used. It provides an important backdrop to the description in Chapter 5 of how this particular research study was conducted and why certain methodological decisions were made.

A number of important considerations need to be taken into account in constructing the theory of an action research methodology. As outlined in this chapter, they are:

6. The evolution of action research
7. The key principles that underpin action research
8. Action research models
9. Action research tools

The evolution of action research

The concept of action research has its origins in the 1940's in the work of social psychologist Kurt Lewin, who developed and applied it over a number of years in a series of community experiments in post Second World War America (Kemmis and McTaggart, 1988:6). The two ideas crucial to Lewin's work were the ideas of group decision and commitment to improvement. He described a

cyclic process which began with examining a general idea followed by planning a course of action, fact finding, analysis, evaluation and making decisions to take further action (Carr and Kemmis, 1986:162).

In Lewin's view it was critical that research should be associated with changing things in their natural situations and that it be participatory and democratic (Poskitt, 1994). In other words, the research site is where the data gathering takes place, where the changes occur and where the people most affected by the decisions are actively involved in making them. He argued for a responsive and flexible approach that took into account the complexity of real social situations. In the deliberate overlapping of action and reflection, he allowed for changes in the action plans as the people involved learned from their own experience (Kemmis and McTaggart, 1988: 8). It was a practical response that enabled groups to live with the 'untidiness' of real experience while, at the same time, striving for concrete improvements.

The value of action research in narrowing the gap between theory and practice and for its potential for galvanising collective action and understanding was recognised by Stephen Corey in 1954 (Kemmis and McTaggart, 1988:6). It was also used by Stenhouse, Elliot and Adelman in the United Kingdom in the 1970's in the Ford Teaching Project, as a means of helping teachers to develop enquiry learning in their classrooms. Stenhouse was the first to raise the notion of teachers-as-researchers. He began a UK tradition that focused on the development of teachers' understanding and theory of classroom practice rather than the theory of action research. The goal was to improve day-to-day practice, not to develop esoteric knowledge. He emphasised that action research, as a method of bringing about curriculum change was dependent on the development of teachers' ability to critique and reflect on practice (Elliot, 1988). Elliot and Adelman advanced this further in the 1980's by focussing on the interrelationships between the roles of the outside researcher and the inside practitioner and the tensions created in learning "how to facilitate reflexive thinking without manipulating and distorting it for our own (the 'outsiders') ends" (Elliot, 1988:32). Schon further developed this notion of

reflection in *The Reflective Practitioner* (1983). He described the reluctance of professionals to subscribe to written theories to guide their actions and the need for them to experiment with their own practice and develop their own expectations, techniques and theories-in-practice.

In contrast, the Australian action research model, based on the work of McTaggart, Carr and Kemmis at Deakin University, challenges the value of focussing on school-based curriculum and pedagogy without consideration of wider social and political relationships and organisation (McTaggart, 1991; Carr and Kemmis, 1986). In their view, action research should be more than practical, it should be emancipatory (Elliot, 1987: 158). Strongly influenced by the critical theory movement and the teachings of Habermas and like philosophers of the 1930's, the Australian theorists advocate a model of action research that requires critical discourse and collaborative community action to expose and challenge aspects of the social and political order that disadvantage and oppress the participants.

There has been considerable and vigorous debate about what qualifies as 'true' action research. As the theoretical basis and methodology of action research evolved, so the definitions changed. In the varying definitions, a different emphasis is placed on the rigour of the inquiry, the use of the research and its impact on the participants.

Elliot (1981: 1) described action research as:

The study of a social situation with a view to improving the quality of action within it.

Watt and Watt (1993: 36):

Action research is a systematic inquiry into collaborative, self-critical communities of teachers which takes place in schools, out of the need to improve educational knowledge and outcomes.

McTaggart (1991: 5)

Action research is concerned simultaneously with improving the work of the individual and collaborating with others to change the institutions and the culture of the societies to which they belong

No particular definition was subscribed to in this study. Rather, the intention was to use action research methodology and then apply a definition that most closely resembled what occurred in practice. In other words, how systematic and rigorous were the four schools? Were they self-critical and collaborative? Were educational knowledge and outcomes improved? Was the process emancipatory? Rather than adopting a strict, and possibly limiting, definition, the parameters of action research can be usefully described in the framework of a set of principles that are subscribed to in almost all the literature.

The key principles that underpin action research

A number of key principles that underpin action research methodology contribute to what Poskitt (1994: 57) calls the 'organisational culture' in which the action will take place.

Firstly, the realities and strengths of working in a **naturalistic setting** will be acknowledged. In other words, there is no attempt to control the variables and the research process is flexible enough to be able to respond to the situation as it unfolds. It is taken for granted that there is no single reality for any group, or groups, of people and that the researcher's values and tacit knowledge are an integral part of the process. Data will be collected and used in ways that allow for these kinds of complexities to emerge. An extension of this principle of naturalism is that the research will also be **responsive to context** (Lomax, 1991). As such, the focus of the research will be centred on problems relevant to the participants, will be context-specific and driven by a practical response to a problem rather than adhering to purely theoretical or academic solutions.

Notions of **democratic participation** are also central to creating an action research culture. Because the intention of action research is to effect change, it is critical that the participants are not only committed to the process but also in control of that process and the consequences of any interventions (Poskitt, 1994: 58). As McTaggart (1991, 40) explains:

A distinctive feature of action research is that those affected by the changes have the primary responsibility for deciding on courses of action and for evaluating the results of the strategies tried out in practice.

Reflexivity is another key principle of an action research culture. It relates to the art of dialectics: the art of asking questions and seeking truth (ibid). Hossack (1997: 10) observed that new learning was only possible when teachers reflected and acted on their tacit knowing. He argued that:

The process of understanding must start from reflection upon one's own experience. The sort of wisdom that is derived entirely from the experience of others is at best impoverished and, at worst, illusory.

Elliot (1993: 3) posits that reflexivity is much more than just reflection involving informal contemplation. It requires critical questioning of one's own action and taken-for-granted assumptions and interpretations. It presupposes a climate where individuals and groups are prepared to engage in honest and open discourse and to hold up their values, beliefs and practices to scrutiny. It presupposes a willingness to take risks and a willingness to challenge.

A great deal has been written about the **collaborative** nature of action research. McTaggart (1991, 5) describes action research as a group activity. Hustler et al (1986: 210) go further and state that collaboration is close to being essential for educational action research. This collaboration extends to including the researchers who are not seen as outside experts but as equal co-workers with the participants, contributing their views and interpretations to the enquiry (Zuber-Skerritt, 1993: 55).

McNiff (ibid) argues that the circumstances of life in a school can militate against collaboration. Teachers face a number of practical difficulties in carrying out rigorous data gathering and finding opportunities to work together on reflection, analysis and planning. Lack of time and funding often mean they do not have access to the kind of support they need, for example, administrative assistance. Integrating research into the normal practice of a busy classroom teacher and not imposing it as an extra demand can be difficult enough, without the added dimension of collaborating with others (Poskitt, 1994: 51; Oja and Smulyan, 1989: 5).

Johnston (1993) also raises concerns about the dynamics of a school staff and the influence they may have on collaboration in the action research process, particularly in relation to power. For example:

[In] school-level changes the social and political context of the school setting becomes more important, as does the level of influence of the teachers attempting to bring about change within the school. The issue of power is important. It is more likely to be successful when senior administrators are involved. They tend to be more skilled and able to use organisational and communication channels (ibid: 24).

Forward (1989, in Poskitt, ibid) suggests that subtle undercurrents among staff may make it difficult, if not impossible, for any individual or group within a school to initiate or sustain collaborative and reflexive action research.

A second set of principles pertains to action research as a valid method of inquiry. They are the principles that help us to measure the dependability and trustworthiness of the piece of action research. Firstly, it must be a **disciplined** and **rigorous** form of enquiry that involves stating problems, formulating hypotheses, gathering data, collating and analysing results and reformulating hypotheses. In other words, it must use the scientific principles of procedure (Poskitt, 1994: 61). Winter (1989, 38) argues that action researchers need to systematically question and test opinions, beliefs and assumptions so that

eventually the understandings and practices are more securely based, or valid, than at the beginning of the study. In this sense, an action research process is **validated in the improvement of practice**. This would suggest, however, that not only would the problem need to be analysed but also, that the action be evaluated.

Validation of action research is also gained through **participant confirmation** (McTaggart: 1991: 6) and by making the study **accessible and examinable in public domain**. As described by Elliot (1993: 3):

The outcome of good action research is not simply the improvement in the quality of teaching for those engaged in it but the systematic articulation in the form of a curriculum design of what this [design] involves and how it might be achieved by other teachers. Good action research does not generate private knowledge for an elite core of teachers. It renders what they have achieved public and open to professional scrutiny.

Another critical factor in the validation of the action research process is the use of an **eclectic range of the methodological tools** in the data gathering process. While predominantly qualitative in nature, the action researcher can also make use of quantitative tools to strengthen the validity of the process and the findings.

Triangulation is another important methodological strategy that can be used for validation purposes. It is not so much a monitoring technique, as a more general method for bringing different kinds of evidence into relationship with each other for the purpose of comparing and contrasting. It will involve collecting data about a particular situation, or part of it, from at least three different perspectives or angles. Contemporary action researchers favour a definition of triangulation which combines the perspectives of the various 'actors' within an action research setting (McKernan, 1997: 184), for instance,

the viewpoints of the students, the teachers, the senior management team and the parents.

Others (Denzin, 1970 in McKernan, *ibid*) extend the notion of triangulation to include a variety of sources. This might involve gathering data from multiple researchers rather than a single researcher; using a number of tools or methodologies; or examining the same problem from multiple perspectives rather than a single theoretical perspective. In comparing different accounts and different sources of data and in that process noting, investigating and even debating the points of difference, agreement and disagreement, triangulation becomes critical to establishing the validity and reliability of action research study.

Action research models

Ultimately, action research is validated in practice. However, a number of different models exist against which that action can be judged to have achieved its objective.

Carr and Kemmis (1986) make the distinction between three types of action research: *technical*, *practical* and *emancipatory* action research. Fitting most closely to Lewin's conception of action research, the aims of *technical* action research are the effectiveness and efficacy of performance through the adoption of systematic, fact finding procedures. The participants are often coopted and rely on the outside 'expert'.

The aims of *practical* action research are also effectiveness and efficacy but the process is allowed to evolve more naturally and is less restricted by the demands of measurement and control (Poskitt, 1994). Rather than just involvement, practical action research focuses on improving participants' understanding and changing practice. The 'outsider' is a process consultant whom Zuber-Skerritt (1993: 47) describes as "playing a Socratic role,

encouraging the participants' cooperation, active participation and self-reflection".

In *emancipatory* action research, the aims are the same as those mentioned above but, in addition, it includes the participants' emancipation from the dictates of tradition, self-deception and coercion and their critique and transformation of bureaucratic systems and organisations. The emphasis is less on technical skills and more on discursive, analytical and conceptual skills with a view to social and political change. Zuber-Skerritt (ibid) sees the facilitator as a process moderator, collaborating and sharing equal responsibility with the participants.

Zuber-Skerritt (ibid) also suggests that, rather than subscribing to only one view of action research as being 'true', there is a natural progressive development from technical through to emancipatory action research. In reality, practitioners might expect to have to move forwards and backwards through, or even between, all three models in order to become proficient and for the process to become natural and embedded. For example, while this particular project most closely relates to the practical model, there were times, especially at the beginning, when the participants' knowledge of action research and new assessment requirements, necessitated following a more technical approach. As time went on, familiarity with the process, greater understanding of the substantive issues, more skills, trust and confidence in the group and in me made it possible to move much closer to the model intended at the outset.

Regardless of the type of model aspired to, the cyclic nature of the action research process applies to all three. First conceptualised by Lewin (1952), it has further been developed by Kolb (1984), Carr and Kemmis (1986), Kemmis and McTaggart (1988), Altrichter et al(1989), Oja and Smulyan (1989), Winter (1989), Elliot (1991), McKernan (1991) and others. It is a spiral of action research cycles consisting of four major phases: *planning*, *acting*, *observing* and *reflecting*. Based on a constructivist view of learning, the assumption is

that learning is experiential and reflective - that is, people can learn and create knowledge on the basis of their concrete experience, through observing and reflecting on that experience, forming abstract concepts and generalisations, and testing the implications of these concepts in new situations. In turn, this will lead to a new concrete experience and, so, the beginning of a new cycle.

Literature on the action research process suggests that the process should be flexible and not adhered to in a rigid and restrictive way. For example, there is no caveat on action researchers entering the cycle at any stage and then following the sequence; the stages can operate in a different order and, if appropriate, simultaneously; and there is no reason why evaluative feedback within and between cycles of action cannot be incorporated (Ebutt in Poskitt, 1994:64).

Johnston (1994: 41) suggests that action research is almost always greeted with enthusiasm by teachers as it is a process that "rings true" with teachers. They readily identify with the process and its terminology and can see direct applications for it in their own teaching. In this study, many of the teachers commented, without prompting, on the way it mirrored the way they teach, assess and evaluate, or how they believe children learn. As stated by McCutcheon and Jung (1990:148):

Action research is characterised as systematic inquiry that is collective, collaborative, self-reflective, critical, and undertaken by the participants of the inquiry. The goals of such research are the understanding of practice and the articulation of a rationale or philosophy of practice in order to improve practice.

In Johnston's view (ibid), this puts into words a process teachers already subscribe to as epitomising good reflective and professional practice. She argues that "they (teachers) are under the impression that if such a process doesn't already characterise their practice, then it clearly ought". Despite its ready acceptance by teachers, Johnston provides evidence that action

research is not something that teachers initiate because, despite its purported flexibility, there are some aspects that are promoted as being essential to the process. Its systematic nature and its adherence to collaboration and critical inquiry are all suggested as not conducive to the practical realities of teaching. She suggests that the very method identified as being a way of closing the gap between theory and practice and between worlds of the academic and the practitioner, in fact, takes on the appearance of positivist traditions in research and teaching: an imposed process, legitimised by non-practitioners to remedy the shortcomings of those at the chalk-face.

While there are obvious difficulties in translating action research into practice in schools, of which the degree of flexibility allowed within the model is one, practitioner skills in data collection and analysis and issues of what counts as reliable and valid data may be of more importance than those related to collaboration and critical inquiry. As this study will demonstrate, many of the teachers involved in the study did not have a repertoire of techniques for data gathering nor the skills for analysing and using that data. Notions of validity and reliability were based on "gut feeling" and "instinct" rather than any precise definitions or judgements.

Action research tools

The purpose of action research is to improve understanding and practice and, as a result, a wide range of tools are deemed appropriate for use. Two primary sources (Kemmis and McTaggart, 1988; McKernan, 1997) have been used to briefly outline these procedures.

Observations are an inescapable component of the action researcher's brief. They can be structured through the use of checklists and ratings or can be less formalised and take the form of anecdotal comments, diary, log or journal keeping and the writing of field notes. Observations by others who have the skills in observational techniques, such as teachers and external consultants, can also be used. Frequent observation, planned and spontaneous

observation and triangulation all reduce the subjectivity of the observations. McKernan (ibid: 62) warns of the disadvantages of observation: difficulty in quantifying the data; the small size of the population observed and the limited generalisability that results; and reactivity where the presence of the researcher may cause the participants to behave uncharacteristically.

Interviews with individuals or **group discussions** with a number of participants can be used to provide more in-depth information or yield questions for further inquiry. They can be either structured or unstructured depending on the nature of the data required.

Workshop and **project meetings** are described by McKernan (ibid) as forums held on a regular basis for gathering project feedback from a number of users, including teachers and students. They allow for evaluative comment, debriefing and problem analysis.

Documents from a variety of sources can provide background information or further written evidence to support or contradict statements made. It can include such things as minutes of meetings, letters, newsletters, school policies and curriculum schemes.

Photographs are another tool for providing documentary evidence. For busy teachers, this is a very easy and cheap way to record events for later analysis.

Questionnaires can be used with any of the stakeholders in a school. They are a quick way of gathering a lot of information from a large group, and provided confidentiality can be guaranteed, allow for honest and frank responses. Careful thought needs to go into the design of any questionnaire. The language used and the length can deter participants from completing the questionnaire and too many open-ended questions can be time-consuming and costly to collate and analyse.

Tape and video recordings can be useful strategies to record lessons, conversations between teachers, interviews and meetings. They can be used by either the researcher or the other participants. Tape recordings do not allow non-verbal behaviour to be recorded and they are not ideal when several people are speaking. However they are useful in interview situations, freeing the researcher to concentrate on the facilitation of the discussion and to note non-verbal responses. Video recordings are ideal for recording complex situations such as interactive classroom situations or at a workshop. Both tools are useful if more than one person is to be involved in the analysis but both are also very time consuming to analyse.

Tests can also be used to triangulate with other data or to evaluate actions taken. They can be informal tests used for formative purposes or more formal, standardised instruments.

While consideration needs to be given to what can realistically be achieved in terms of the data to be gathered (Elliot, 1991: 83), it is crucial to remember that evidence of validity is shown in the rigour of the data collection, the appropriateness of the tools used and the responsiveness of the data collection to the context in which it is occurring.

Summary

Action research has its origins in the work of Lewin and Stenhouse and has since been extended and refined by researchers world-wide. The literature reveals that it is a complex and involved methodology, involving the direct participation of the stakeholders, enhancement of understanding and a commitment to the improvement of practice. A number of different models of action research have evolved over time, each with a different emphasis on the balance of research and action, theory and practice. There is general agreement on the broad principles that govern its use and these include naturalistic inquiry, reflexivity, collaboration and responsiveness to the context in which the research is occurring.

The way in which the theory of action research translates into practice in the four ABLE schools is the focus of the next chapter.

CHAPTER FIVE

THE METHODOLOGY IN PRACTICE

'If you want to understand what a science is, you should look in the first instance not at its theories or its findings, and certainly not what its apologists say about it; you should look at what the practitioners of it are doing' (Geertz, 1973: 5).

Introduction

This action research study began in August, 1995 when ERDC⁹ successfully negotiated an Assessment for Better Learning (ABLE) Contract with the Ministry of Education. The Contract required the ERDC Team to work with sixteen schools in the Auckland area to provide a programme of development in assessment. This chapter outlines the organisation of that Contract and how four of those schools were selected to participate in the research. It also explains the different roles and responsibilities of the teachers directly involved in the project and those only indirectly involved. The practical and ethical issues that arose during the study are discussed along with a description of how the data were analysed. The limitations of the research are recorded so that the reader can interpret the findings with these in mind.

As this chapter parallels the previous chapter, the headings under which these topics will be discussed are as follows:

1. Origins of the action research process in the four ABLE schools
2. The action research model
3. Tools used to gather the data
4. How action research principles were reflected in the study
5. Data analysis

⁹ Educational Research and Development Centre. This is a self-funding centre within the Massey University College of Education. The author is based at ERDC at the Albany Campus of Massey University.

6. Ethical issues
7. Methodological limitations

Origins of the action research process in the four ABLE schools

In the Education Gazettes¹⁰ on September 1st and 15th, 1995, all schools (primary, intermediate or secondary) in the Auckland region were invited to participate in the ABLE Contract. A covering letter, an overview of the Contract organisation and programme and an Expression of Interest form (see Appendix B) were posted out to each school that responded to the advertisement. Expressions of Interest were received from 139 schools out of a possible 450 (31%).

The breakdown of the types of schools that applied was:

Secondary	23
Intermediate	16
Primary	97
Special schools ¹¹	3
Total	139

Schools were then selected for a ballot using the following criteria:

- situated in the Auckland metropolitan area (rather than in the wider Auckland region)
- consultation carried out with **all** staff in deciding to apply for the Contract
- had unsuccessfully applied to participate in a 1995 ABLE Contract

¹⁰ The Education Gazette is an official Ministry of Education publication that is produced fortnightly and distributed, free of charge, to all schools. It is used by the Ministry to disseminate official notices and by educational organisations and schools to advertise such things as vacancies, professional development opportunities and resources.

¹¹ Special schools include hospital schools and schools for the physically and intellectually disabled.

This limited the number of schools to fifty. The schools were then divided up into primaries (including intermediates) and secondaries, then into geographic areas and a ballot was conducted to select the final sixteen schools - four secondary schools and twelve primary schools.

One of the methodological strategies was to set up small groups of schools within the Contract that would meet five times over the year for discussion, sharing and critique. The twelve primary schools were then divided into three groups with some diversity being achieved in each by including a mix of :

- school types
- decile rankings¹²
- school sizes

At that point, all the schools were asked to confirm their participation in the contract programme and each of the Principals of the schools in the most diverse of the three groups was asked to consider taking part in the research. This group was selected because they represented such a broad range of school types, year level combinations, deciles, and roll numbers, allowing for any differences attributable to those variables to emerge.

The table below provides the statistics for each school in relation to the selection criteria and demonstrates their diversity.

Table 1

School	Classifi- cation	Type	Year levels	Decile	Roll No.	No. of Teachers
A	State	Intermediate	Years 7 - 8	3	312	14
B	State	Primary	Years 1 - 6	2	391	17
C	State	Primary	Years 1 - 6	5	408	16
D	Private	Full Primary	Years 1 - 8	-	302	13

¹² New Zealand state schools are ranked on a decile scale of one to ten. It indicates the school's socio-economic status, calculated by assessing the socio-economic position of a sample of the students in the school. The closer the decile is to ten, the higher the socio-economic status of the school. Schools with a decile ranking between one and four receive additional funding from the government.

Each of the Principals was given a copy of the research proposal. The methodology and the implications for the schools were fully discussed and they were given the opportunity to ask questions and raise any concerns. All of them responded very positively to the proposal. Two of them were currently involved in post-graduate studies themselves and were especially enthusiastic. All four of them commented on the benefits of a methodology that allowed for reciprocity.

I like the idea of there being some direct spin-offs for the school. Quite often we get asked to fill out forms for this research or that and you never know what happens to it all. This way, we contribute and we also get the feedback. I like it! I'm sure that will help "sell" the idea to staff too. (Principal)

Because all four schools had new teachers joining the staff in the new school year, it was decided that both the principal and I would discuss the research project at a full staff meeting to coincide with my initial visit to each school in early February. That staff meeting provided me with an opportunity to meet all the teachers, to discuss both the contract programme and the research and the links between them and to clarify the respective roles of the lead team, the teachers and my role as a participant researcher. In all cases, the decision to take part in the research was made after I had left the meeting. By mid-February, all four schools had agreed, in writing, to be part of the research.

All teachers in all four schools participated in the study. For purely practical reasons, however, each of the four schools appointed a lead team that coordinated and facilitated the process in their school. Because the teams were small, it was easier for the schools to find and fund relieving teachers to release the team members for regular day-time meetings. At the time, Auckland was experiencing an acute teacher shortage and many schools were struggling to find classroom teachers, let alone relieving teachers. This was particularly so for the two schools in the study located in South Auckland.

While all staff participated in the cycle of data collection, analysis and reflection, planning and acting, the lead teams were more actively and directly involved in the process. Sometimes, this simply meant 'doing the leg work', where the team would carry out decisions made by the whole staff. For example, one staff decided to analyse the different assessment techniques teachers were using and asked the lead team to design a form for staff to complete. At other times, it meant that the lead team made the decisions. In one school, the lead team analysed samples of portfolios across all classes, identified a number of professional development needs and then decided on a series of actions to address them. It was also the lead teams who met five times over the course of the year to share and critique ideas and to discuss and reflect on the work they were doing. As one teacher commented:

I've learned a lot about assessment and I have a bit of a handle on action research but there's no way I feel I know it like the others (the Lead Team) do. (Teacher, School A)

Each lead team comprised three or four members. Schools made their own decisions about who would be on the team and the status and experience of the teachers varied from school to school. However, there was a predominance of senior management team members (principals, deputy and associate principals) and senior teachers (invariably leaders of a team of teachers). In three of the schools, the principal was a fully participating member of the team. While much professional development literature supports the notion of the active involvement of the principal (Stewart and Prebble, 1993; Hopkins, Ainscow and West, 1994) and the need for them to collaborate closely with teachers (Stoll and Fink, 1996), Hossack (1997) warns against principals unduly influencing the direction of the group. She advocates that the rules, roles and responsibilities of the group, and individuals within the group, are clarified to avoid this happening. Although Hossack wrote this after the study began, these issues were discussed fully with groups at their first meeting with the researcher. Close examination of the transcripts indicates that none of the principals exercised undue control over the way the groups

functioned or the actions they took. The table below describes the composition of each team.

The action research model

An action research model (see Table 2 p.66) was prepared prior to the study beginning and was shared and discussed with all the staff when the researcher initially visited each school. It was also elaborated on when all sixteen schools combined for the first professional development workshop.

As described in the previous chapter¹³, many action research models have been developed, but all are based on a cyclic process involving some form of planning, acting, observing and reflecting. The model used in this study attempted to provide a structured, easy-to-follow start in order to give the schools confidence in the action research process.

It also aimed to accurately describe the tasks involved, in language that would not be ambiguous. For example, many teachers equate observation with the narrow interpretation of watching students at work in a classroom. Therefore, because it would more accurately describe what the teachers would be doing, observation was substituted by data gathering.

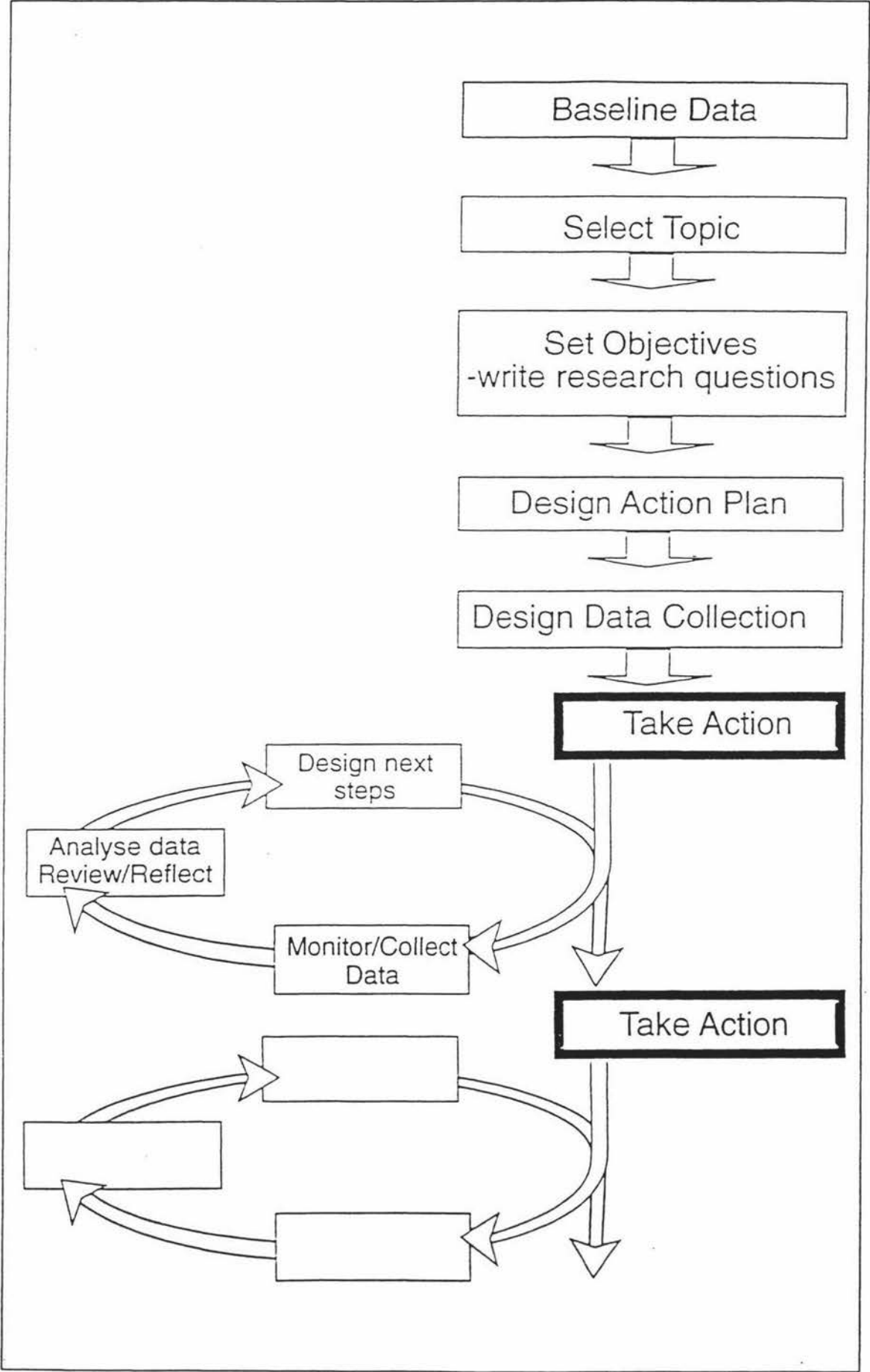
There was also a desire to use words that more closely mirrored the language teachers use when they talk about assessment. This provided an opportunity to reinforce the similarities between an assessment 'cycle' and an action research process.

Tools used to gather the data

Once the four schools had agreed to participate in the research, the action research process began with the collection of some baseline data. Each of the teachers in the schools completed a simple post-box **questionnaire** (see

¹³ See p. 53

Table 2: The model used to guide the process in the four schools



Appendix C). In all cases, they were completed at staff meetings or at a syndicate meeting. Each teacher was provided with an envelope in which to put the questionnaire and these were handed in to the respective lead teams for analysis.

There were a number of advantages in beginning with such a simple data collection tool: it provided a quick, safe and confidential way of getting baseline data from staff; its simplicity meant the lead team could analyse it themselves with minimum help; and it enabled each school to get started immediately on their first action research cycle. The intended impact for the schools was to allow them to gain further confidence in the action research process. For the writer, the intention was to provide a useful baseline against which subsequent developments and changes could be measured.

There were three strands to the Contract programme that were woven into the action research process. Those three strands were:

- school-based visits
- development group workshops
- the Help Desk

In addition to the initial visit to the school where the research project was discussed with staff, a minimum of five **school-based visits** were built into the programme. These visits were organised around meetings with the lead teams, all of which were taped and later transcribed. The agenda of these meetings was based on the needs of the schools at the time and where they were in the action research cycle of data collection, data analysis and reflection, planning and acting. All schools kept a record from meeting to meeting of what they had done and what they planned to do (see Appendix D). The format for these records varied from school to school and, depending on the notetaker, could change from meeting to meeting within a school. Regardless of the format, they were referred to by the researcher and the school personnel as **action plans**. All four schools struggled from time to time with keeping these notes

and planners up-to-date and in making them available to, and discussing them with, the rest of the staff.

The school-based visits also provided other opportunities to collect data. These included **observations** of staff meetings and syndicate meetings that focussed on assessment, **informal discussions** with the principals and with teachers and **informal visits** to classrooms. At one school, the researcher attended a teacher-only day. On all of these occasions, detailed field notes were kept.

Six times during the year the schools met for what were called **development group workshops**. The first of these was held with all sixteen schools on the Contract. It occurred just after the initial visits were made to the schools and after the baseline data had been collected. The purpose of this meeting was to give all the schools an update on recent assessment theory and research, to clarify and discuss the legislative requirements and Ministry guidelines relating to assessment, to talk further about the action research process and for each school to plan the initial steps they would take.

The remaining five development group workshops involved just the four schools. The first two meetings were for full days, subsequent meetings were for a morning each, and all were designed to provide a forum for the schools to work together. They were important opportunities to share ideas and information, swap resources, problem-solve, debate processes being used and options for resolving blocks to change, and to critique and challenge others' work. The group decided they would prepare an agenda for each of the meetings and that this would be organised in advance. A blank agenda sheet was faxed to each school at least a week prior to the meeting date and items or issues for discussion were recorded (see Appendix E). Sometimes, the faxes were used to request examples and resource information from the other schools. The researcher then combined the ideas and circulated a finalised agenda to each of the schools. This gave the schools control over the content

of the workshops and how their time was spent. It also provided some structure to the sessions and ensured that everyone came along well prepared.

In research terms, both the lead team meetings and the development group workshops were, in effect, group discussions. In order to keep the distinction between the two, however, they will continue to be referred to as lead team meetings and development group workshops. In both cases, tapes were made of the discussions and were later transcribed.

Another feature of the contract that provided an invaluable opportunity to collect data was **the Help Desk**. The schools were able to phone or fax the researcher at any time and a log was kept for each school, detailing the kind of help required. For the schools, it served the purpose of providing them with support between the school-based meetings and the workshops. It formalised the ongoing contact and gave them ready permission to clarify issues, discuss concerns and to seek affirmation and critique. This facility was used extensively by two of the schools and to a lesser extent by the other two. It was most commonly used for:

- accessing articles and background information
- accessing samples eg assessment strategies, report forms
- checking a data collection process eg a staff questionnaire, questions for a review of a portfolio system
- seeking help with data analysis eg interview data, a parent questionnaire
- critique of documentation eg a policy, a letter to parents on the purpose of profiles
- accessing the name of a school or schools to visit or get help from

- advice on team or staff dynamics
- accessing general information eg the validity and reliability of PATs

Close to the end of the research programme, **individual interviews** of two teachers from each of the schools were conducted by the researcher. Because much of the research had focussed on the work of the lead teams or on the staff working as a whole or in syndicates of at least three teachers, the lead teams made the decision to collect more in-depth data from individual teachers via semi-structured interviews carried out by the researcher. Each school selected two classroom teachers who had not been on the lead team, were representative of different year levels or syndicates at the school, had been teaching at the school for no less than eighteen months, knew the school and the evolution of its assessment practices well and who were respected practitioners. These interviews were not taped but detailed notes were taken as the interviewees talked.

Finally, **an end-of-contract evaluation** was carried out (see Appendix F). The lead teams worked in groups at the last workshop to complete a series of open-ended questions designed to guide them into planning for the following year. This information was later typed up used by the researcher to triangulate data collected earlier in the study.

Table 3 on page 72 summarises the information described in this section.

How action research principles were reflected in the study

In Chapter 4, a number of essential principles were defined as being essential to the implementation of an action research methodology. The principles were grouped under two broad headings - those that serve to create an action research culture and those that contribute to action research as a valid mode of inquiry. This section takes those two broad headings and the principles that fit with each and describes how they were reflected in the methodology of this

Table 3

Techniques used to gather and triangulate the data	Examples of when techniques were used	How data were recorded
Group discussions	Lead Team Meetings	Taped, then transcribed; field notes
Workshop, Project meetings	Development Group Workshops; Teacher only day (1 school only)	Taped, then transcribed; field notes Field notes
Interviews	Informal interviews of Principals and teachers	Field notes
	Semi-structured interviews of 2 teachers in each school	Taped then transcribed
Observations	School and syndicate meetings	Field notes
	In classrooms	Field notes
Help Desk		Log
Questionnaire		Individual post-box questionnaire End-of-contract lead team evaluations
Document Analysis	Analysis of individual school <ul style="list-style-type: none"> - Action Plans - Policies - Initial school wide assessment recording and reporting documents - Examples of teacher assessment documents - Documentation developed during the year 	Notes

particular study. (See Table 4 on P.72). Some of the procedures were determined at the start of the study and others were incorporated as they emerged from the action.

Table 4

Principles of Action Research		Procedures used to reflect the Principles
Related to an action research culture	Naturalistic	<ul style="list-style-type: none"> - each school <ul style="list-style-type: none"> • set their own goals • decided on action appropriate for their situation • was in control of their own action plans
	Responsive to context	<ul style="list-style-type: none"> - adopting then adapting the Contract process - the Help Desk was used as a vehicle for the researcher to respond to schools requests for support and guidance - action plans were set up at each lead team meeting; they were constantly added to and changed - the researcher participated in extra meetings/group discussions e.g. a TOD, staff meetings, syndicate meetings
	Reflexive	<ul style="list-style-type: none"> - combined workshops with the four schools were used to critique each others' work and to generate options for resolving problems - the Help Desk was used to fax ideas/ documents to the researcher for critique - lead team meetings were used to debate and plan - regular Milestone Reports for the Ministry prompted researcher reflection
	Democratic and participatory	<ul style="list-style-type: none"> - schools: <ul style="list-style-type: none"> • self-selected to participate in the Contract and the research • selected their own lead team • set their own goals • participated in a cycle of collecting data, analysing data, reflecting, planning and acting.
	Collaborative	<ul style="list-style-type: none"> - the lead teams in each school met on a regular basis - involvement of staff other than the lead team was mandated in the Contract - action plans were designed to be shared and discussed with staff - resources were shared between the schools
Related to action research as a valid method of enquiry	Disciplined and rigorous	<ul style="list-style-type: none"> - the lead team and/or staff used action plans to guide their process - agendas were prepared in advance of all meetings with input from the researcher and the schools - Milestone Reports were prepared each term for the Ministry and were used as opportunities to collate and analyse data
	Eclectic (in terms of techniques used)	<ul style="list-style-type: none"> - tools used: <ul style="list-style-type: none"> • group discussions • workshop meetings • interviews • observations • Help Desk • questionnaires • document analysis
	Triangulated data	<ul style="list-style-type: none"> - a number of different techniques were used to collect data - data were collected from four schools allowing comparisons and contrasts to be made - the schools' data, as well as the researcher's data, were used in the analysis

This table illustrates that each of the key action research principles was accommodated in the methodology in a number of ways. In practice, however, it was more difficult to achieve the desired result than the table would indicate. A critique of how well the principles were adhered to and the factors that influenced both the researcher's and the schools' performance in this regard, is outlined in the next chapter.

Data analysis

Data analysis occurred throughout the research process. The focus for the schools was the data they were collecting themselves but, at the same time, the researcher was feeding back her own analysis of transcriptions of taped group discussions, fieldnotes, interview data and documents. There were times when the analysis was shared with only one school. This was particularly the case when analysing the specific process occurring in each school. At other times, it was appropriate that the analysis be taken back to the collective schools. The Development Group workshops were an ideal forum for this to occur.

The data analysis began with the school post-box questionnaires. These were analysed by the researcher and each of the Lead Teams and then used, in consultation with staff, to identify each school's initial research focus. As the project progressed, all transcriptions and field notes pages were given a number. This page number was then repeated down the margin at the start of each paragraph. Codes were also assigned to each paragraph to indicate the school or the interviewee and the occasion eg lead team meeting, Development Group workshop, informal discussion (see Appendix G).

As the transcripts and notes were read and re-read, key ideas began to emerge. A word or two to describe these were then written down the right-hand side of each page. In some instances, ideas within the larger ideas were also identified. As more data were gathered, emerging themes were identified and numerous analytic memos compiled to relate similar ideas and concepts. The next step was to photocopy, cut up, then sort the numbered and coded

transcripts into the themes. The frequency and importance of the data determined the significance of a theme and whether it was discounted or included in a large wall chart onto which the themes and the sub-themes were organised. In the final phase of the analysis, the individual quotations and notes were then sorted into a logical sequence and taped onto sheets of paper.

In the final individual interviews with the two teachers from each school, the researcher raised some of the key ideas that had emerged from the data and asked each of the participants to comment. These comments and opinions were integrated into the analysis both directly as interview notes and indirectly, by assisting in the understanding of the data as the study was written up.

Ethical issues

One of the researcher's responsibilities as a researcher was to ensure informed consent was sought from each of the schools (Burgess, 1989: 64). As mentioned earlier, it was decided to discuss the research at a staff meeting at the beginning of the new year because all four schools were expecting new teachers to join the staff. In all cases, the decision to join in the research was made after I had left the meeting and the agreement to participate confirmed in writing and signed by the principals. Informed consent was also sought from the eight teachers who participated in the eight interviews. Prior to the interview taking place, the researcher discussed the purpose of the interview with each teacher and how the data would be used. Each participant was given the opportunity to decline to take part.

The research also raised issues of confidentiality. It was very important that the lead team meetings and the development group workshops were forums where people felt safe enough to be completely honest and where they could raise sensitive and difficult issues. Because the schools were going to be involved in making changes it was inevitable that not all staff would be able to make those changes smoothly and that there would be resistance. As Nias (1993: 150) suggests:

Change is both threatening and inherently conflictual. It is threatening because it involves the loss of structures which give our lives meaning and so it carries with it the fear of disintegration.

Evans (1996) takes this further when he describes the ambivalence associated with change. On the one hand, change raises hope because it offers growth and progress - but it also stirs fear because it challenges competence and power, creates confusion and conflict, and risks the loss of continuity and meaning.

Both the in-school meetings and the workshops needed to be forums where the lead team members could talk about difficulties they, themselves, were having, the difficulties others on their staff might be experiencing and the impact of these issues on the process. I also needed to know that sensitive information that I gave to the principals, or the lead teams was kept in confidence. This meant that the staff involved needed to feel safe both with me and with each other.

This was addressed in several ways. At the first meetings of both the lead teams in each school and the development group meetings, the purpose of the groups was discussed and the roles and responsibilities of the members and the researcher clarified. This included the need for confidentiality. It was agreed that any personal issues or issues related to a particular staff member or groups of staff that were discussed in those forums, would not be discussed outside the meetings. The groups became self-regulating and it was not uncommon, particularly at the beginning of the year, for one of the members of a group to remind the others about the code of confidentiality. No breaches of the code were raised during the year and as the trust grew, particularly amongst the four lead teams who had not previously worked together, confidentiality was taken for granted.

It has been very important to be able to speak our minds at these workshops. It took a couple of meetings for us to get to know each other

but I've appreciated being able to talk up front about some of the hassles we had. On more than one occasion, it has been crucial. (Principal)

The issue of confidentiality was also raised in feeding back the outcomes of the interviews with the eight teachers. Because there were only two teachers involved from each school, it was going to be very easy to identify who had said what, especially since they all taught in different areas of the school. It was agreed that a summary of the feedback from all eight teachers would be given to all the lead teams. It would then be up to the lead teams to extrapolate data of particular relevance for their context.

The official nature of this contract was an important incentive to the schools in maintaining commitment to the programme for the whole year and an obligation to make significant progress. The Contract provided the schools with enough money to release the teachers in the lead teams to participate in the school-based meetings and the workshops with the other schools. There was also a stated expectation in their contract that they would commit an equal amount of money from their own professional development budgets which meant that the lead team and other teachers could carry out additional work between those meetings and workshops. This made it easier for the schools to put more time and energy into the project and, as a result, they were able to make greater gains more quickly than might have been expected if they had been doing this work without that help.

Because the Ministry was funding the programme, there was the danger that they would also be in a position to control the content and outcomes of the programme and, by implication, the research. Although the Ministry's permission was sought for the writer to undertake the research, no alterations were made at that point to the programme and no restraints were placed on the writing up of the thesis. Their only request was that they get a copy of the final document.

Initially, the schools were cynical about the Ministry's motives and suspected they had a specific agenda, even though teachers were not always able to articulate what they thought that agenda might be. While it was a prerequisite that the work of the schools would be based on the National Curriculum Statements, the ERDC programme was very clearly designed to be needs-based and it was accepted that each school would decide what they would work on and the direction that development would take. In fact, the action research methodology helped to demonstrate to the schools quite clearly, that they were in control of the work they did and the direction they were taking.

As the facilitator of the Contract, I was directly involved in all the "formal" interactions that made up the Contract programme - the initial workshop for all sixteen schools, the five Cluster workshops and the four school-based meetings. I participated fully in each of those meetings and talked, laughed, asked questions and listened like the others.

The contradictions in the 'insider' and 'outsider' relationships in the action research process have been well documented and debated (Elliot: 1998). Poskitt (1994:82) describes issues of power, knowledge, access to and use of information and goal incompatibility as being four areas that are fraught with difficulty for the 'outsider'. There was considerable evidence to suggest that, in the schools' view, the positive outcomes outweighed the potential detractors of 'outsider' involvement in the project. Encouraging reflexivity, keeping the schools focussed and motivated, providing new insights and expertise in research skills were all mentioned by the participants in the end-of-Contract evaluations as being important to the changes made in both their thinking and their practices .

For me, the biggest challenges related to issues of knowledge and power. Much of this stemmed from the schools' expectations of being involved in a development Contract sponsored by the Ministry of Education. They were buying into a process with which they associated the provision of expert advice and direction. It was a constant challenge, particularly at the beginning, to

remain sufficiently conscious of the process to ensure that my beliefs and knowledge did not dominate or, in any way, diminish the capacity of the 'insiders' to reflect, critique and negotiate their own changes.

My role in the group discussions changed over the course of the year. The transcripts clearly showed that my input was greater at the beginning and lessened dramatically as the year progressed. At the school-based meetings, the teachers' getting to know and trust me and vice versa, and becoming familiar with the Contract organisation and expectations were critical to the shift in my role. In the workshops, getting to know the others in the group, becoming more familiar with the different school settings and the work in which each school was engaged were more critical. To begin with, many of the questions were directed towards me, both in relation to the Contract process, including the action research, and to assessment itself. The transcripts of the later meetings show that, as the teachers began to trust their own practice and become 'experts', a lot more interaction occurred between the members of the group with very little input from me.

Data gathering began almost immediately in the research process. It occurred on two levels, firstly by the schools as they worked through their own action research cycles and secondly, by me, as I worked with the lead teams both individually and as a group.

The individual school data gathering was always analysed by the lead teams or by some lead team members with other staff coopted for the purpose or by the whole staff. Who conducted the analysis depended on who was available when the analysis needed to take place, whether the analysis coincided with a lead team meeting or whether particular staff or all staff needed to be involved. The data were presented in a variety of ways. For example, the results were sometimes recorded in paragraphs or under bullet points and at other times, graphs and tables were produced. On a number of occasions, data were analysed and no written record was kept as it was fed immediately into the next

set of actions the group was going to take. Records of all these analyses and plans were kept and have been incorporated into the thesis.

The schools received ongoing feedback from me as I read the literature, analysed the lead team discussions, interpreted the issues raised by the schools at their development group workshops and carried out the teacher interviews. They were generally presented as oral summaries either at the end of a meeting or at the beginning of the next meeting. After two of the workshops, the participants asked for a written summary and these were provided. Although ERDC was required under the terms of the contract to write Milestone reports for the Ministry of Education they often contained confidential material and included references to schools who were in the contract but were not part of the research. For these reasons, they were not shared with the schools and used as a reflexive tool. These Milestone Reports, however, served as a record for my own level of the research process.

When the groups were given feedback, the transcripts recorded their responses. These responses provided a useful means of validating the data. It is also important to note that, although I was giving the schools ongoing feedback, they did not always act on it. There were a number of examples in the transcripts where feedback did not result in any in-depth or useful analysis. Sometimes they had, what they considered, more important issues to discuss and they would move onto something that was more immediate and urgent for them. This was accepted as part of the process and I deemed it to be in the best interests of both the group and the thesis that it be allowed to occur. There were times when the group would remember the feedback at a later date or an issue would resurface allowing me to reiterate an earlier message.

Methodological limitations

The framework of the contract programme was designed by the ERDC Team before the decision was made to incorporate the research. The lead teams, school-based meetings and development group workshop were not only non-

negotiable components of the programme for the schools, but also for the researcher. It could be argued that this framework detracted from the purity of the action research process. Writers such as Carr and Kemmis (1986), Kemmis and McTaggart (1988) and Oja and Smulyan (1989) make it clear that ownership and control of the process by participants is an important and distinctive aspect of action research. Ideally, the schools would have collaborated with the researcher to decide on the best structures for working together as a group of four schools and for facilitating the development within their schools. In practice, those structures were already in place and the schools' ownership and control began at the next level - making decisions about the direction they would take and how they would go about it.

The Contract methodology also dictated that much of the researcher's time be spent in group discussions and in workshop and project meetings. This meant that less time was able to be spent in classroom observations and conducting individual interviews with teachers. This was compounded by the areas selected by the schools to be the focus of their research. Many of them were school-wide issues and, as a result, more data were gathered and more themes emerged to inform school-wide assessment systems and processes than individual teacher practices. This also meant that the themes which emerged from the study had a more summative than formative emphasis.

A further limitation was the length of the study. Both the content of the study (assessment), and the process (action research), were new areas of learning for almost all of the schools and the participants. As one teacher remarked:

We were on a vertical learning curve for most of the year.

In the researcher's experience (1997 ABLE Contract: Final Report, 15) and supported by other contractors involved in Ministry contracts (McAlpine et al (1998); Willis and Bourke, 1998), a year is insufficient time for schools to institutionalise changes. Dalin and Rolff (1993: 145) calculate that projects at the classroom level which really change the role and behaviour of teachers and

students can take from four to five years to be implemented throughout the entire school. The schools in this study certainly went through the initiating phase and began the process of implementation. None of the schools reached the third phase that Miles et al (in Dalin and Rolff, *ibid*) describe as institutionalisation. On this basis, this thesis does not document the resolution of some of the important issues it raises.

Summary

The action research methodology used in this study indicates some modifications to that described in the *Methodology in Theory* chapter. Responsiveness to the context required that the research had to 'fit' within the contract methodology and this placed some constraints on the participants' initial involvement in setting up the research and in making decisions about some of the data gathering tools that were used. The Help Desk was an important innovation that formalised the process of the participants gaining support, guidance and feedback as they worked their way through their action research cycles. It also gave the researcher ongoing insights into the processes each of the schools went through.

Initially the lead was taken by the outsider but, over time, as the teachers' knowledge of assessment grew and they gained a greater understanding of the research process, more responsibility was able to be handed over to them. Ideas that were formerly taken for granted were challenged; concepts that were previously just 'gut feelings' about what constituted good practice were clarified and articulated; alternative ways of operating were debated. The processes and the modifications to practice that resulted are the subject of the next chapter.

CHAPTER 6

EMERGENT THEMES

'Assessment is an opportunity for teachers and students to reflect, question, plan, teach, study and learn' (Earl and Le Mahieu, 1997: 158).

Introduction

The initial baseline questionnaire carried out by the schools determined that five key aspects of assessment were the focus for development by the schools over the course of the year. These aspects (or contexts) were: the place of assessment in the teaching and learning cycle, developing school-wide recording systems, reporting to parents, developing standards or benchmark portfolios and student self-assessment. This does not mean that other aspects of assessment were not explored by individuals, groups, or even the whole staff, but they were the focus of most of the schools' time and energy. All of the schools worked on at least two of these aspects and they were all discussed and debated collectively at the Development Group workshops. Quite clearly, the schools' focus was on summative assessment. At the time, it was a disappointment to the researcher that only one school chose to address an aspect of formative assessment but because it was crucial that the schools take control and ownership of the process, it was essential the schools' own data guide the developments.

As indicated in the chapter *Methodology in Action*, the developments that occurred in the four schools revealed some emergent themes. These themes were evident in the transcripts of meetings, in the field notes, the interview data and in the data and documentation generated by the schools themselves. A repeated process of sorting, analysing, and rechecking both during the study

and at its conclusion, determined six major themes as the most predominant. These themes were:

1. Understanding assessment theory
2. Understanding the curriculum
3. Planning and design
4. Control of the assessment process
5. Honesty
6. Consistency

Understanding assessment theory

Despite the fact that all of the schools had spent some time the previous year evaluating and modifying their assessment systems, a number of them had not articulated the concepts behind or the fundamental reasons for making the changes. Sometimes, the basis for the changes they were making had little to do with assessment. One school discarded the Primary Progress Records because:

Some of the schools are getting rid of the coloured cards (the Primary Progress Records) so we thought we'd better take a look at it. It's such a competitive market around here that you can't afford not to keep up with the play. (Principal)

There was also a desire to be in charge of their own systems rather than be constrained by those imposed by the Ministry. They wanted to collect the assessment information they believed was relevant to meeting the needs of their particular teachers, students and parents.

Often, the new or modified documents were an add-on to a string of other records that teachers were using. Little rationalisation was taking place because teachers were not being asked to examine their own individual practices in light of the school-wide changes being made. For these schools,

over-duplication of information and complaints of assessment overload were clearly expressed in the initial data gathering exercise.

For many of the teachers, the Contract provided an important opportunity to think in-depth about their school's assessment systems and their own classroom practices in the context of assessment theory. This was reflected in the quotes taken from the evaluations at the end of the first workshop and from subsequent discussions with individual teachers:

It was the dawning of understanding! It had never fitted together like that before. (Senior Teacher)

For the first time I got some answers that made sense to me. It gave me the confidence to question some of the things (about assessment) that come out from the Ministry. (Teacher)

The time we spent sorting out the underlying principles (of assessment) was great. It just clarified a whole lot of things. (Teacher)

One of the barriers to teachers' understanding assessment theory is that none of the schools had an established research culture. In fact, very few of the teachers, other than the Principals and those who had studied or were studying for a diploma or a degree, had engaged in any academic or research-based reading.

If it's longer than two pages, forget it. They want information that can help them tomorrow when they front up to their class. It's much easier to get them reading something that is practical - a 'how to' article. (Principal)

The Help Desk provided an opportunity for the Lead Teams to access articles and books related to assessment. Two of the schools devised interactive activities to introduce some articles at meetings with staff. Their initial fears

that some staff would disparage having to read academic material or that the exercise might discourage interest in, or provide an excuse for not, making any changes in assessment, proved unfounded.

Another barrier was the terminology associated with assessment. There were times when terminology was dismissed as jargon and this made it difficult for the concepts to be discussed. Sometimes, teachers did not have the words to describe what they meant. Even some of the basic terminology like formative, summative, moderation and criterion-referenced was an obstacle for some participants. There were instances where different meanings were attached to the same words.

We had these two staff meetings and had some interesting debates. One teacher got quite emotional and was quite upset about the whole concept. And what we worked out... it was all to do with the semantics. The minute I said I think we've got the wrong words here, let's use other words, we actually got there. People were talking levels and they got really confused because we didn't identify what we were talking about clearly enough. (Senior Teacher)

In some instances, different terminology was used to describe the same concept and when the 'experts' engage in this, the confusion is compounded. The lack of a common discourse with which to discuss assessment stopped some teachers from contributing to discussions, created confusion, slowed down the pace of change in a number of situations and, at worst, prevented a small group of teachers from taking ownership of the process occurring in their school.

The baseline questionnaire asked the teachers in the four schools to identify, what they considered to be, the key purposes of assessment. The table below identifies the teachers' priorities.

Table 5

	A	B	C	D	Totals
Key purposes of assessment	N=9	N=15	N=15	N=10	49
Planning and Evaluating programmes (evaluation)	2	12	9	4	27
Identifying students' needs (diagnostic; formative)	7	12	4	4	27
Tracking student progress (summative; accountability)	3	4	5	4	16
Reporting to parents (reporting)	2	1	2	4	9
Placement of students (placement)	-	1	3	-	4
Providing feedback for students (formative)	-	1	1	-	2

The letters A-D refer to the schools. The totals in each column do not add to 49 as respondents could record more than one key purpose. All the numbers indicate raw scores as the sample was too small to convert to percentages. 49 represents an 81% return from across the four schools.

The aggregated data show that the diagnostic/formative and evaluative purposes were of greatest and equal importance to the teachers. The diagnostic and formative functions have been combined because many of the teachers did not distinguish between the two. Diagnostic assessment was often described as identifying students' needs. Giving students feedback is also formative assessment but is listed here separately because, in all the responses, only two teachers mentioned this as a key purpose of assessment. For many of them, the locus of control for assessment is still firmly with the teacher. Even parents do not feature as a priority in the assessment process.

Tracking students' progress was generally described in summative terms and was invariably linked with accountability issues. When the four schools entered the Contract, all of them had already abandoned the Primary Progress Records or were using a combination of them and their own records. One school had introduced student portfolios and two others were in the process of developing a system of individual student records which they later called profiles¹⁴. The fourth school had just spent a year trialing what they called a Student Discussion Document. This was an individual student profile that also served as a discussion document for use with students and parents at the joint conferences held with teachers at the end of each term. Keeping individual student records was a priority for these schools. If they were going to discard the Primary Progress Records, there was an urgent need to have something in its place that would meet the requirements of the NAGS and stand up to the scrutiny of the ERO. Given the public nature of the ERO reports and their ability to influence the perceptions of a school of both parents and the wider education community, it was not surprising that three of the schools chose as one of their primary goals to improve the ways in which they monitored student progress over time.

Placement referred to the placing of students in classes on their arrival in the school. Many of the teachers did not place a lot of emphasis on or give credence to the judgements made by previous teachers, especially those from other schools. For some, even the judgements of those from within their own school were not trusted or valued. As one teacher commented at a staff meeting:

I just don't rely on what other teachers write down. Half the time, it doesn't tell you anything or it's not accurate enough. Anyway, I like to give the children a clean start, especially if it's someone from another

¹⁴ A profile is an individualised record that contains both quantitative and qualitative summative data (generally summative) to describe the achievements of each particular student. It is a cumulative record that is added to throughout the child's time at the school. It does not contain samples of the student's work and is primarily a document for teachers.

class (in the school) and they've come with a bit of a reputation. Sometimes I don't look at their records until..like...maybe until six weeks after they've arrived. Unless they create havoc on the first day, then I'll look it up! (Teacher)

While this teacher was challenged by others at the meeting, there were several teachers who agreed with and supported her position.

Until the teachers at each of the schools had spent some time discussing, debating and clarifying the ideas and concepts on which assessment is based and had gained some confidence in using the terminology, it was difficult for them to make judgements about the kind of assessment systems they wanted to design and implement. It was also difficult for teachers to challenge each other about their individual practices so they, in turn, could make decisions about any other individual records they would keep. Up until then, the basis of their judgements had often been: *that's the way I've always done it*.

The other area critical to individual teachers and schools making appropriate decisions about their assessment systems and ensuring that would work for them was their understanding of the curriculum, its content and its organisation. It is to this area that the discussion now turns.

Understanding the curriculum and the expectations of the Ministry

From the outset, there were many occasions when teachers referred to their knowledge of the curriculum as a determinant of their assessment practices. If they were going to make valid assessments of students' progress and achievement, a prerequisite was a good understanding of the curriculum content.

Assessing Science is hard because I'm not certain about it myself. It helps if I'm clear in my own mind about what I'm doing and what I'm looking for. (Teacher)

It's hard to assess things I don't do well. Like Music. I don't teach it well so I can't assess it. In fact, I always hope it will go away! What is really scary is having to turn around and label a child when you're not too sure about it yourself. (Teacher)

A number of teachers also referred to the ease with which they assess when they have a good understanding of the developmental stages in knowledge and skills acquisition or where their knowledge of the content is such that they can help students make the links between what, for them, may be isolated ideas or understandings.

It's easy for me to assess in reading because I know what I'm looking for. I'm trained in Reading Recovery and it's second nature for me to make judgements about what the children need at any particular time. (Teacher)

In Maths, I know what the building blocks are so it's not a problem to describe exactly what the children can do and what they need to work on next. I can talk to the children about it and help them to make the connections too. (Teacher)

In all four schools, the Lead Teams quickly came to the realisation that another critical curriculum prerequisite to good assessment practice was a working knowledge of the curriculum statements. Once the teachers knew how each document was organised, had understood the objectives and had identified the school's priorities within those objectives, they were in a more confident position to make decisions about what would be assessed and how to go about it.

Getting to know the Document was essential. We had to be confident using the document first and know where things were., then we could concentrate on the planning. Then came the assessment. (Senior Teacher)

If you haven't got a good grasp of the document or if you haven't got your scheme sorted out, then teachers find it really hard to do the assessment. It makes their planning easier which frees them up to spend more time dealing with assessment....they can see where it (assessment) fits in. (Principal)

There was strong evidence to suggest that when teachers are confident with the content of the curriculum, can make the links between the different ideas and understandings and are familiar with the curriculum statements, they are more likely to plan and assess in an integrated and coordinated way across the curriculum.

You have to be aware of the assessment possibilities within any unit of study. So if you're doing magnets in Science and the children draw a poster then that's visual language and you can assess both. It makes it so much easier if you don't compartmentalise the learning areas (Senior Teacher)

To begin with, there was a great deal of debate about the organisation of the curriculum into levels. The teachers expressed suspicion of the Ministry motives in using levels, especially as the official assessment guidelines (Ministry of Education, 1994: 34) advocated that summary records produced by schools should include a statement of the numbers of students working at each level in the different curriculum areas. The guidelines went on to say that, as part of the report to parents, a statement should be included giving the level or levels at which each student is working in each strand of each curriculum area (ibid: 38). What concerned the schools was not the organisation of the curriculum into levels but whether the levels were going to be used to compare

schools and to norm-reference individual student achievement. Fears were allayed considerably when they were told by the researcher that, at a Ministry workshop for contractors working with schools on ABLE contracts across the country, a senior Ministry official announced:

We knew they (the particular references to levels) were unreasonable but they were included on the insistence of the Minister. It was the only way we could get the document (the guidelines) published.

Most teachers were only using the levels as a guide for establishing the section of the curriculum document that applied to their particular class. Although it must be remembered that some of the documents were still very new to the teachers, some teachers were unaware of the objectives that applied to other class levels. Some others found it useful to *know where they* (the students) *have come from and where they are going to*, but teachers still planned from the one curriculum level, even in the senior classes where the range of abilities was often more pronounced.

The most heated discussion on levels centred around the decision of the intermediate school to include a levels continuum on their report forms. This decision was taken in response to parents' request for some quantifiable data to support their children's applications to get into particular and desirable secondary schools. The intermediate teachers were sometimes nervous about making accurate judgements without any standards against which to make those judgements and, quite clearly, in the absence of those standards, were able to manipulate the data to their own ends.

At mid-year, I don't mark on the upside. What I want to do is show some movement. I want to show that they (the students) have made some progress. (Teacher)

Some of the other lead team teachers had forgotten that this was commonplace for more experienced teachers who were familiar with the P and A Registers and the Progress Cards.

The other tension that surfaced in relation to levels was whether to follow the levels as set out in the curriculum documents or whether to remain with the developmental levels that many schools had been following for some time. This particularly applied to levels in reading and writing and arose when designing individual students' profile documents and when compiling benchmark portfolios. What is important to note is that, in making their decisions, none of the schools opted for profiling or benchmarking in class levels, a clear indication that the teachers in these schools were focussed on a constructivist approach to learning.

One of the most important statements on assessment that all the teachers in the schools read was the Education Gazette article published in March 1995 (Ministry of Education, 1995:1). Even though the article had been published 9 months prior to the Contract starting, many of the teachers in the schools had not read it or had not fully comprehended its implications. When the article was discussed with the four schools, the key messages they drew from the article about the Ministry's view of assessment were: assessment procedures should be non-intrusive for students and focussed on promoting learning; schools should exercise professional judgement in providing a balanced curriculum and select achievement objectives for assessment that reflect that balance (in other words you cannot teach everything and assess everything in the curriculum); and it is not possible to report the level reached by each child in every strand, in every curriculum area. The emphasis on meeting students' needs and exercising professional judgement appeared to place decision-making firmly back into the hands of the schools. It validated what the schools were wanting to do. They no longer felt there was 'one right way', the Ministry's way, and this gave them permission to take more risks and work on developing systems they felt were appropriate for them, their students and their parents.

The next section outlines the ways in which the teachers used their new understanding about the importance and place of assessment in the learning and teaching cycle to change the way they planned for assessment. It also describes their reluctance to make use of the outcomes of those assessments to better inform their planning. As well, it highlights the lack of confidence a number of the teachers had in their validity of the assessment tasks they designed.

Planning and design

Over the year, there were many discussions in the school-based meetings and in the workshops about where assessment fits into the planning cycle. Each of the Lead Teams built on or developed new strategies to encourage teachers to see assessment as an integral part of the learning and teaching process and to see it, both in formative and summative terms, as an ongoing process. There were few instances in the four schools where assessment was not seen as part of the planning process or where teachers would not engage in some summative assessment as part of each major unit of work.

All of the teams of teachers in the schools were involved in joint planning of some kind. In some instances joint planning included deciding on a topic or some common objectives and brainstorming some possible activities. For other schools or teams it meant planning in fine detail, including making decisions about common assessment tasks.

Some of the planning was based on particular topics and was driven by the context. For example, the starting point for the planning might be volcanoes or Pacific festivals or magnets. The basis for the choice of topic might be tradition, availability of resources, the teachers' preferences or a combination of any of the three. When the planning was context-driven, the curriculum objectives were made to 'fit' the chosen topic. For some of the other teams, even within the same school, the starting point for planning was the objectives from which a suitable context was selected. Some of the schools developed

two-year overviews that served as a guide for teachers' planning and the way these were organised (by context, by objective, or both) would quite frequently determine how planning was carried out.

The data tentatively suggest that context-driven planners tended to plan their unit around a series of interesting activities and then decide which activity they would use as an assessment task. When teachers planned from objectives, they were more likely than the context-driven planners, to plan specific assessment tasks to measure specific outcomes.

An important consideration when designing an assessment task in our syndicate is whether it meets the requirements of the objective we want to measure. It has to be real and relevant and within the context (of the unit) and definitely not an add-on. (Senior Teacher)

Despite the fact that considerable effort was often put into joint planning, there was little evidence of teachers evaluating the assessment outcomes together and then using those outcomes to inform the planning of their next unit.

If we plan a unit together, we do the same assessments but we don't share the results. We don't think about it, we're on to the next thing. (Teacher)

We miss out on the step of going back - to see if we are consistent. I think we're still stuck in the planning stage. We don't even go back to see if we treated the assessment in the same way or to see if the children did well overall. We don't evaluate the results as a team. (Teacher)

In the original baseline questionnaire, teachers were asked to list the assessment techniques they felt very knowledgeable about and confident in using. The table below describes their responses.

Table 6

Assessment Techniques	N=49
Standardised tests e.g. PATs, Running Records	48
Observations including conferencing	27
Written tests	21
Conferencing	14
Check lists	12
Self-assessment	8
Portfolios	5
Very few/none	2
Peer assessment	1
Gut feelings	1
Marking children's work	1

The total does not add to 49 as respondents could record more than one key purpose. All the numbers indicate raw scores as the sample was too small to convert to percentages. 49 represents an 81% return from across the four schools

Of the 49 respondents, almost all said they were knowledgeable and confident in using standardised tests, although it is interesting to note that the actual tests were referred to eg PATs¹⁵ or Running Records¹⁶ rather than the generic technique ie standardised tests. They also made reference to the recording tools, such as checklisting rather than describing the assessment technique that was used to generate the data. Only one participant mentioned peer assessment and, other than a small group listing portfolios, no other mention was made of performance assessment. The other important omission was

¹⁵ PAT is an acronym for Progressive Achievement Test. There are PATs available for Year 4 through to Year 9 students in Reading Comprehension and Vocabulary, Listening Skills, Mathematics and a range of Study Skills. They are New Zealand-designed and norm-referenced.

¹⁶ A Running Record is a standardised performance reading test used to determine a student's reading age.

reference to informal formative assessment. Conferencing with students was mentioned but this suggested a more formal assessment situation. Only one reference was made to giving students feedback. In discussions with teachers, they often said they are constantly assessing students in the classroom and that they do it intuitively. It would appear that formative assessment does not count when they think more formally about assessment and that, in some way, it is different from assessment.

In Stiggin's (1991: 535) view, teachers who have the assessment design skills sufficient to measure outcomes effectively are able to:

- understand the meaning of high and low quality assessment and are able to apply that knowledge to various measures of students' achievement
- ask two key questions about all assessments of student achievement: what does this assessment tell students about the outcomes we value? What is likely to be the effect of this assessment on the students?
- seek and use assessments that communicate clear, specific and rich definitions of achievement that is valued
- know the importance of using an assessment method that will reflect a precisely defined achievement target
- realise the importance of sampling a performance fully
- are aware of extraneous factors that can interfere with assessment results
- know when the results are in a form they can understand and use.

Three considerations emerged as priorities for the teachers in the study schools. When they selected which activity they would use as an assessment task or designed a specific assessment task, they based their selection or design on the importance of the knowledge or skill being assessed, manageability and student appeal. Its level of importance was decided by its 'newness', whether it was identified in the profile for assessment, whether the teachers judged it to be important or because they knew it was a particular need of the students at that time. Manageability was defined as being not too difficult to mark or too time consuming or within the teachers' comfort zone. A small group maintained that the test had to be 'fun' for the students, in assessment terms, have face validity. Validity was rarely, if ever, discussed unless it was raised by the researcher. The teachers understood the concept to be important to assessment but, by and large, were unable to articulate a description of validity or its application to designing assessment tasks. Common responses included statements like the following:

It's a gut feeling, it's an automatic thing - I hope! (Teacher).

You just do it (planning assessment tasks) on instinct and you get to be very accurate. You learn to trust your judgement. (Senior Teacher)

Some said they were unsure if their assessment tasks were 'good' but trusted the experience and judgements of the team with whom they planned. Some said that if they got the information they wanted, then they knew it had been a well-designed task but they tended to make those judgements retrospectively. In 1989, Crooks' review highlighted that teachers have little or no formal training in educational measurement techniques and that there is a need for improved skills in observation and non-test means of assessment. The data in this section suggest that these findings remain valid.

Despite this, the perception of the teachers is that they are meeting the needs of the children more effectively because their teaching and learning programmes are more closely aligned to objectives and because their

assessments are more ongoing and systematic. Some are also saying that they are teaching less but teaching it better.

We used to have 2-week units but we're expanding them to three weeks now. We're more conscious of assessment and more aware of learning. Teaching to objectives and looking at the outcomes is slowing down the process.

There were also indications of teachers' willingness to involve stakeholders, other than teachers, in the assessment process.

Control

For many teachers, assessment was regarded as largely belonging to their professional domain. Summative assessment data was seen by many as privileged information that only teachers were qualified to fully understand and which was very likely to be misinterpreted by parents. For many of the teachers on the Contract, giving students an opportunity to contribute to the assessment process in a formalised way was also not common-place. For example, even in those classrooms where self-assessment was included in the programme, it was often regarded as a fun activity, something to use as an end-of-unit activity rather than a prerequisite to further learning and the development of meta-cognitive skills. During the course of the study, there was evidence of an increase in teachers' willingness to directly involve other stakeholders in the assessment process. For some, there was a growing understanding of the need for learners to have ownership of the learning process and for other significant adults (other than the classroom teacher) to support that learning. These shifts in the locus of control will now be discussed.

Involving parents

In the previous year, one of the schools had already become disillusioned with their written report system. There was a tradition in the school that, once a

term, the school was unofficially closed and parents and children came to school for a 15 minute interview with the child's teacher. The focus for this group conference was a written report for which teachers prepared a written comment in each learning area. They were proud of this inclusive practice but were frustrated with the apparent lack of appreciation of parents of their efforts and their repeated calls for more specific information. The decision was made to design an individual student profile that would be the official record of each student's progress through the school, as well as a replacement for the report for the first three terms. They were encouraged by the positive response of the parents.

What it gave them was what their child had achieved and what standard of achievement they had reached. But what impressed them the most was they could see what they (their child) had to do next. That was new information for them. And because we were showing them all of our records, they were also amazed at the amount of information we were collecting. (Senior Teacher).

This inspired another school to rethink its reporting programme. In the past, the reporting system had been dominated by tradition, the way things had always been done, and what had best suited the school's or the teachers' timetables. The pattern of a 'Meet-the-Teacher' evening, mid-year and end-of year reports and such things as dates for the school production and inter-school sports exchanges had long been paramount in their decision-making. Using feedback from staff and the outcomes of a parent survey, the Lead Team drew up a reporting schedule, process and format that they hoped would more closely meet the needs of parents.

It included:

- using the individual student profile for discussions with parents mid-way through terms 2 and 3
- renaming the parent interviews as conferences

- sending home a list of 2 or 3 items the teachers would like to discuss at the conference and inviting parents to make their own list - this was returned with the parents' preferred times for the conference
- at the conclusion of the conference each teacher recorded 2 or 3 actions that would result from the discussion and these were followed up at the next meeting or, in term 4, in the end-of-year report
- several teachers volunteered to involve students in the second round of conferences
- the 'Meet-The-Teacher' evening, which was described by many parents as "*a waste of time*", was replaced by an early March report to the parents in the next year. This report would include information on their child's adjustment to the new class, their relationships with their classmates and the teacher, and their work habits.

All of the innovations worked to a greater or lesser extent and the trial was to continue into 1997. The staff were disappointed with the number of parents who returned the 'agenda' slips (approximately a third in each class) but the numbers increased the second time round. Many teachers found it difficult to quickly negotiate the action plan in the time given and, as each conference was already 15 minutes long, the teachers were reluctant to extend it further. By the time the second conferences were held, just over half the plans had been actioned and the teachers were surprised at how seriously they were being viewed by the parents and some of the students. In the classes where parents were encouraged to bring their children, not all did so, but teachers reported that these conferences were highly successful and wanted them to continue. The lead team was convinced that, if this practice was to succeed in the long term, they would have to make compulsory.

In both the schools discussed in this section, profiles were used for reporting to parents. In both instances, the result was a heightened awareness by parents

of the depth of knowledge teachers bring to the teaching and learning process and, in the views of some teachers, an improvement in the professional profile of the teachers in their schools.

There was a general consensus by the end of the year, at least among the lead teams, that parents have a right to information about their child's achievement and have a right to have it presented to them in an understandable form or, at least, the opportunity to have it explained and to ask questions.

There was also a growing recognition that, like teachers and students, parents also needed time to adjust to having more opportunities for inclusion in the assessment process and to get used to participating in different ways from the past. The school that had a portfolio system reserved 6 to 8 blank lines on the last page for parents to record their comments. They were disappointed with the lack of parent comment on some and the very negative comments on others. They held an Open Night to talk about the purpose of the portfolios and how they were intended to be used and then, after feedback from one group of teachers, added some unfinished sentences to the blank lines to guide the parents in making their comments - *something I like about the work you have done is...*; *something I think you could work on is...*; *something that surprised me is...* By the end of the year, the parents' contribution to the portfolios had improved markedly.

Only one of the schools was considering computerising its recording and reporting system. They did not want to be coerced into using a commercial package as all of those investigated did not match the particular format or content they wanted. A senior member of staff had the necessary computer expertise to enable the school to develop its own package that would replicate the manual system they had developed over the year. Even though they had the expertise, acquiring the necessary hardware for all staff to have access to a computer could only be achieved in the medium term. They intended to begin by trialing a computer package with the three senior class teachers in the following year. Insufficient computers for staff use were the main reason the

other schools were not computerising their records. For them, computers for students' use was cited as a higher priority.

Involving Students

There were a number of examples of teachers endeavouring to involve students more directly in the assessment process. These included student self-assessment, involving students in setting unit objectives, involvement in conferences with parents and seeking student feedback when evaluating programmes. While none of these examples is new, they were departures from usual practice for these schools. What they learned was crucial to their adopting more inclusive assessment practices and shifting away from teacher-dominated teaching and learning processes.

One school was struggling with student goal-setting. It was intended to be an important part of their portfolio system but some teachers were expressing dissatisfaction with the vagueness of some of the students' goals and were sure that nothing was changing as a result. They were unsure that goal setting warranted the time and effort being put into formulating them. Some were describing it as 'a fad'. At a staff meeting, teachers examined a random sample of the portfolios from their class. They agreed that one of the reasons the students were not doing this task well was that the teachers, themselves, were not skilled at setting specific and achievable goals. They were unfamiliar with the developmental progressions in attaining this particular skill and were unsure of what they should be expecting from the students. A further complication was that, at the beginning of the year, the teachers were asking students to set academic goals when they did not know the students well enough to guide them. Using a process that they were using in their own teaching teams, the whole school adopted a thematic approach to student goal-setting. Their focus for the first term was work habits and behaviour, in the second term they moved to literacy and numeracy, in term three they broadened the scope to include other learning areas and in the term 4, focussed on preparation for secondary school (year 8) and the school camp

(year 7). The second critical change required the students to describe how they were going to achieve their goals and to check in with their teacher twice a term to report on their progress.

In contrast to setting term goals, one syndicate at another school trialed setting weekly goals with their children. Structuring the task around something they could achieve by Friday helped the students to focus on small and manageable tasks.

Teachers across the four schools were experimenting with involving students in setting objectives. The school using portfolios as their school-wide assessment system included this as a requirement for all work that was put into the portfolios and the objectives. As teachers became familiar with including a discussion of the objectives with students, two other changes occurred. For some teachers, the practice transferred to other lessons where work was not necessarily going to be included in the portfolio and two of them (from the lead team) experimented with devising the objectives with the students, using their ideas and their language. The notion of rubrics was foreign to these teachers. Because they were experimenting with self-assessment, many for the first time, the extra task of preparing criteria against which students could measure their success seemed too daunting for the lead teams to contemplate. While some of them could see the sense in designing rubrics for or with their students and were excited by what they were reading, the number and extent of the changes they were already making and the number of assessment priorities already competing for their time, precluded them from trialing further developments in this area until the following year¹⁷.

As already indicated in this section, two of the schools involved students in their parent/teacher conferences. In the school where it was already accepted practice, teachers came into the school knowing that this was expected of them, had formalised guidelines to refer to and had a pool of other teachers'

¹⁷ Schools in the 1997 ABLE Contract were given the option of continuing with the Contract for a second year. Three of the study schools did so.

experiences to draw upon. In the second school, the most confident teachers decided to trial it and, while they had few concerns, their colleagues worried most about how they would deal with sensitive information in front of children, whether it would stop them from giving honest messages, whether the children would misbehave or be a distraction, how they would handle parents bringing along the whole family and whether they had the skills to deal with a group. In their feedback to staff, the teachers were overwhelmingly positive about the inclusion of the students. Some students did change their behaviour because their parents were there, but the teachers accepted that what they observed was an important part of their learning about the families' relationships.

They also decided that the students deserve to have the expectations of the conference and their role clearly articulated. One teacher suggested that the best training for the conferences was providing more opportunities for student self-evaluation and regular formalised conferencing with teachers to discuss their work. The teachers reported that the children were able to be very honest about their strengths and their weaknesses and they were able to give one or two examples where the students' honesty had prompted more open discussion or, at least, more honesty about what action needed to be taken. It is difficult to single out one factor as being most influential in the changes these teachers experienced in their new approach to conferencing, but they did note that having the students involved kept the focus very much on the student, their most important needs and how these could be addressed.

There were no generalities. The kid just said, "I've got problems with my spelling. I'm lazy because I don't like going back. I hate looking up words". He said it much more plainly than I could have said it and we just got on with dealing with it. (Senior Teacher)

Except for one school which carried out group discussions with their exiting year 6 students, none of the schools engaged in student evaluation of programmes or teachers. For a number of the teachers, there was resistance to both receiving and giving honest feedback.

Honesty

The focus on assessment raised many issues for the schools and teachers in relation to honesty. It was most often discussed when the schools were discussing sharing data and documents with parents or when they were reviewing their reporting practices. It was also raised by teachers in two of the schools in relation to students.

Some teachers saw the outcomes of a student's achievement as a direct reflection of their ability as a teacher and very often felt they did not have the evidence to support the statements they were making or the grades they were awarding.

Before, it (the report) had to be a good reflection of me and quite often I found myself on the back foot trying to justify what I had done. (Teacher)

I am just saying that we get frightened about being honest. We'd rather not put the bad news in (the report). There might be 101 reasons why Ashley's reading age hasn't gone up. One might be the teacher, it might be everything else that's happening at home. But we'd rather label the child or the parents than ourselves. (Senior Teacher)

Some teachers deliberately acted as gatekeepers and protectors, based on their own perceptions of what was 'good' or 'bad' for parents.

I sometimes share the school entry assessment information with parents. I showed it to one parent because she would follow it up. In fact I show it to most of them if they are in and out of the classroom. But I wouldn't if their child hadn't done well. That would be detrimental. It wouldn't help them or the child at all. (Senior Teacher)

Some teachers were fearful of being too honest and open with students about their achievements, or lack of, and, in particular, disclosing their level of

achievement. This exchange between two senior teachers illustrates the dilemma.

Teacher 1: If a child has a low self-esteem and is working at Level 2 and he opens up his profile or his report and says, "I'm at Level 2 or I've got a reading age of 9, I'm dumb", that's just reinforcing it.

Teacher 2: He knows that already, he's in Form 2. My children in Std 1 know that they are not the best at reading. It's what you are doing about it that matters. Anyway, you'll often find that they are very good at something else.

Particularly in the junior areas of the school in the study, teachers were very anxious about the children being aware of being tested and the impact this might have on their relationship with the child. They wanted to be perceived as facilitators rather than adjudicators or judges. They wanted to keep any testing situation low-key and preferred that students and parents only make comparisons with the student's previous performances rather than with the performances of other children.

I don't want them to think I'm examining what they do, that I'm testing them. They're my children and my relationship isn't a testing one, it's a helping one. And I'm reluctant to promote competition. I want them to grade against themselves, not each other. The parents too. They have to accept that making comparisons with other children isn't helpful.
(Junior Teacher)

This contrasts starkly with the results of a parents' survey carried out by one of the schools. After asking a number of quite specific questions, the parents were asked to record any other improvements they would like the school to make. Many of the responses related to receiving accurate and specific information about their child's progress and for early notification if there is a problem or a concern about their learning.

Written reports should tell us honestly what our child is achieving. We don't want this positive stuff that we know to be untrue. (Parent)

I want to know how well our child is doing in relation to where they should be. (Parent)

It is immoral and disempowering to mislead us over our child's progress. By not being honest you block us from assisting in any way. (Parent)

The teachers debated the ethics of not giving honest feedback and concluded that not being honest with children and parents can be a barrier to their learning. For some of the teachers, being honest was associated with being negative and they acknowledged that they are sometimes unsure how to deliver honest messages while at the same time, making them positive and motivating. For some, it was acceptable to be honest with words and grades but unacceptable to give parents, let alone students, norm-referenced information.

There are early indications that some of the changes they are making to their assessment practices are giving the teachers the confidence to be more honest, at least with parents. Their new recording systems are requiring them to be more specific about what the students are setting out to learn and what, at the end of the unit, they can and cannot do. While a number of them acknowledge that this puts the spotlight on teachers and makes them more openly accountable, it also provides the evidence they need to justify the judgements they have made about a student's performance.

In one school, there is also data to suggest that a staff decision to make their student portfolios into working documents rather than showcases, has persuaded teachers to be more honest, has helped them encourage students to be more honest and increased the likelihood of teachers informing parents earlier if there are issues causing them concern.

The portfolios keep you honest because they contain the hard evidence. You have to put it up front. Under the old P and A system¹⁸ you could make guesses and then write flowery reports. I think we're coming out of that dishonest phase. (Senior teacher)

If you're not honest, it sets kids up for disaster and misinforms the parents. You're better to deal with parent grief sooner rather than later and I prefer to do it before the portfolio goes home. (Teacher)

Another major theme that emerged from the data was consistency. It was identified early in the study as an issue for many of the teachers in all four schools and was often referred to by the lead teams and the teachers on subsequent occasions. It was a theme that cut across a number of aspects of assessment.

Consistency

The need for many of the teachers for consistent, school-wide practices was first identified in the baseline surveys and was evident in the goals the schools set for themselves and in their planning. It translated into a number of practices the schools put into place as a result of that planning and into their expectations of teachers.

One area where teachers wanted consistency was in curriculum planning. Two of the schools already had syndicate overviews in place and a third school spent the second half of the year preparing them for implementation in the following year. Typically, these spanned a two-year period and matched the year levels around which the national curriculum is organised. As explained in an earlier section, not all of the overviews were linked to the level objectives in the curriculum; some were based on topics or contexts. The purpose of the overviews had quite practical origins. It was an effective way of utilising, rationalising and sharing resources; it simplified teachers' planning and

¹⁸ Progress and Achievement Register. See P 11

prevented the same contexts being repeated in other classes at different levels. While the overviews were not intended to be a straitjacket, and teachers gave examples of how they adjusted their programmes to accommodate special events and students' interests, the overall outcome was a structured and consistent programme across the school.

The redesigning of their school-wide recording systems also resulted in greater consistency, regardless of the format (profile or portfolio). While the schools had all avoided "*the uncritical adoption of mechanistic schedules*" (Ministry of Education, 1995:2), they devised recording systems which required teachers to record the results of particular tests or to assess students against certain objectives and/or skills and standards. Although they arrived at the outcome in different ways, each school went through a process of identifying specific priorities for summative assessment. The next steps were to identify how the data would be collected and then to devise the best way to record it. This often resulted in a combination of quantitative and qualitative recording formats. In the case of the school with portfolios, consistency was reflected in the requirement to include certain information with each assessment task, for example, the objectives of the unit, the criteria for marking, a student self-assessment and detailed formative feedback from the teacher. The consistency achieved by having a school-wide recording system, going through a process of ensuring that all their other documentation had an express purpose and that data was not being duplicated in any way were important steps in ensuring that the assessment process was manageable for teachers.

Before the Contract came along, I was overwhelmed with all the paperwork I had to do. I had my roll book, the special notes that I kept in the back of my workplan, the Progress Cards, some other sheets we [the Junior Syndicate] had added in along the way, syndicate records. It was alarming. I just wanted to give up [teaching]. We all knew there was a lot of doubling up but nobody seemed to know what to do about it. I certainly didn't. I just thought it was the new system [Ministry requirements]. (Teacher)

A further example of the schools' seeking consistency was the development of benchmark documents. These are sometimes referred to as benchmark or standards portfolios and contain samples of students' work, dated, annotated and with names removed, to demonstrate different levels of achievement. Typically the levels are based on developmental progressions (eg early, emergent, fluent) or on the curriculum levels and can be used by teachers to determine where individual students, a group or a class are placed on a continuum of achievement in a particular learning area. Two of the schools worked on these during the course of the study and this generated a great deal of interest with the other two schools. There were two key reasons for establishing the benchmarks: to demonstrate the next steps in learning as a guide for teachers, students and parents and to achieve a higher level of consistency in marking and making decisions about a student's level of achievement. Senior staff in particular, wanted teachers to be making similar judgements about similar work. They had particular concerns about supporting beginning teachers, teachers from overseas and those they described as 'weak', to maintain standards consistent with what teachers had discerned to be appropriate for particular levels at their school and to make assessments consistent with their colleagues. They wanted parents to be able to trust that they were getting accurate and internally consistent messages about their child's achievement levels and their progress. Until they began to work through the process with their staff, some of the lead team members, echoed later by some of the teachers, resented that the Ministry was not providing national exemplars. In an evaluation of the process at the end of the year, the feedback from the teachers was telling:

The process was really valuable to the staff because it was the first time that we'd all got together in the same room and talked about actual (students') work. So much came out of it. (Senior Teacher)

Everyone got drawn into it, even the resistant ones (teachers). It was the best professional development we've ever done. We had to do it as a whole staff because the practical manipulation showed how the

process worked. Making the comments (the annotations) was the hardest. The discussion got so animated! It was unbelievable!
(Teacher)

It was a very long and drawn out process but it's good to have a consistent standard. (Teacher)

At that time, it was the intention of the schools to develop benchmarks in a few key areas but they recognised that the process would take considerable time. One school talked about linking it with their curriculum review process. In the views of the two schools that embarked on this initiative, quite clearly the benefits of developing their own standards would achieve greater consistency than imposing standards from 'the outside'.

After extensive consultation with staff, one school devised and mandated a school-wide, long-term planning sheet. This also prompted considerable debate at one of the last meetings of the Lead Teams. It raised the issue of teacher autonomy and its tension with school-wide or syndicate consistency.

Teachers are happy to use this? I regard planning as a very personal and individual thing and when you start to standardise planning, you are saying to teachers, "Do it our way"! (Senior staff member)

There was a general consensus that as a result of the introduction of the new national curriculum, the requirements of the National Education Guidelines and the developments in assessment practices, there was now less autonomy for teachers than was previously the case. They also acknowledged that some of the calls for consistency had been motivated by the government's demands, and those of parents, for greater teacher accountability. Nonetheless, the members of the lead teams decided, and this was supported by almost all the teachers interviewed towards the conclusion of the research, that developing consistent practices across their schools had increased their ability to identify and address students' learning needs.

Setting out the expectations (in the school's guidelines for student portfolios) was so important. Now we know what to expect of ourselves and others in the team. I was always wondering if mine (the students' portfolios) were good enough and if they were doing the job they were meant to. (Senior Teacher)

Now, I'm a lot more specific and the assessment tasks are a lot more pertinent. The individual profiles make me think about what each child has achieved. (Senior Teacher)

The teachers were also more clear about what they had to teach and the importance of measuring, either informally or formally, the learning outcomes. For many of them, it signalled a shift away from basing the measurement of outcomes on intuition to one based on evidence and written records.

Going from the objective to the task to the profile has made the outcome more definite. Before I used to assume. There was a lot of guesswork. (Teacher)

The profile has given me more things to think about (compared with the Primary Progress Records). Before, Maths was Maths and I thought about what the children were doing but I didn't really focus on the different aspects in relation to each individual child. The profile forces you to focus and makes me think a lot more. I'm picking up more information like she can do this, she's good at that, but not good at this and I'm seeing patterns. (Teacher)

Summary

The understanding of assessment theory and a working knowledge of the content and organisation of the curriculum were two crucial and connecting threads that ran through the other themes identified in the data: planning and design, inclusion, honesty and consistency. None of the themes, however, can be seen in isolation and all represented important departures from previous

practice. As the teachers' knowledge increased and they began to challenge previously-held assumptions, they contemplated changes, then accommodated new ways of operating, in relation to their attitudes towards assessment; their role, and the roles of others, in the process; and alternatives to current systems. For some, it began to transform their learning and teaching programmes. The following chapter discusses the implications of the changes to practice that occurred and of the themes that emerged as those changes were initiated and then implemented.

CHAPTER 7

INSIGHTS AND CONCLUSIONS

'Our mission as educators is to help every child become a more active, engaged, committed and skilled learner, not just for a test, but for a lifetime' (James Bellanca, If Minds Matter, Vol. 2).

Introduction

According to Eisner (1993: 224), correct policy formulation does not improve what happens in classrooms. Policies are easy to formulate and even easier to mandate. For him, the problem is one of practice and in his view:

Good teaching and substantive curricula cannot be mandated; they have to be grown.

This final chapter draws together the insights and learnings of both the researcher and the teachers in the schools. It discusses and summarises how the teachers in the four schools 'grew' their assessment practices and examines these in light of the current expectations and policy requirements of the Ministry. It will be argued that, despite the language of the Ministry policies and their insinuation that learning and the quality of the learning programmes can be improved through accountability rather than through effectively integrating assessment with learning, the schools have maintained their autonomy in this regard. They have been able to mould the requirements to integrate with or build on their own individual and particular pedagogical philosophies and practices.

The section on assessment literacy of teachers raises some major concerns that have been highlighted in the literature over the last decade (Crooks, 1989; Eisner, 1993; Black and Wiliam, 1998). Many of these have not been fully

addressed in the four study schools and there is no reason to assume that the situation is any different in other similar schools. Until teachers are able to create assessment procedures that have educational validity and have some measure of confidence in their judgements about the internal validity of those assessments, it is difficult to imagine how schools can continue to resist calls for nationally mandated testing. Some of the schools made significant moves to involve other stakeholders in the assessment process, particularly parents, and this whole issue of teachers' preparedness to share the control of assessment information and of the reporting process is summarised.

Some key elements of the methodological process will also be discussed and an evaluation made of the impact of these elements on the changes made by the schools. It will also suggest some critical prerequisites to empowering teachers as researchers. The chapter ends with possibilities and recommendations for research into assessment in the future.

These insights and conclusions are grouped under five major headings:

1. Autonomy and accountability
2. Assessment partnerships
3. Assessment literacy
4. Methodological insights
5. Future research

Autonomy and accountability

At the beginning of the study, there was an expectation on the part of the teachers, that developments in assessment would dominate the content and delivery of the curriculum. They were fearful of the prescriptive and mechanistic checklists they had seen emanating from other schools and, because of the publicity these systems were getting in educational circles, were suspicious that this was, in some way, being endorsed by the Ministry and ERO. All three schools with profiling systems steered away from the checklists.

Their systems were based on identification of assessment priorities. Sometimes the basis for these priorities was the new curriculum and sometimes they used assessment schedules that had worked well for them in the past. This happened most commonly with reading and writing where developmental progressions, often originating in the junior school, were used.

This research suggests that the recording requirements, either summative (the profiles) or formative (the portfolios) were not dominating the learning and teaching process. By the end of the year, many of the teachers commented that the new systems, and the process they went through to devise them, had enhanced their ability to more accurately focus on students' needs and to define outcomes with a greater degree of certainty. In accordance with the views of Khattri et al (1995), that the more teachers learn about what students think and how they learn, the more likely it is to affect their pedagogy, there was ample evidence that this occurred with the teachers in this study. Furthermore, there was tentative data to indicate that the new systems were impacting most significantly on a particular group of teachers. This group were identified by Gipps et al (1995) and Hill (1997) as intuitives or head-note assessors whose assessment practices are characterised by guesswork. Gipps et al (1995) described this type of assessor as highly resistant to assessing against statements of attainment. These are the teachers who, in the past, used to 'call up their memory' to arrive at summative assessments, generally in the middle and at the end of the year. Because there is more pressure on them to use their recording documents in an ongoing way for team planning, for recording specific assessment outcomes and, increasingly for reporting, there was evidence of shift away from their previous ad hoc practices. The indications are, in this study at least, that the same radical shifts did not occur with the other types of assessors: the integrated/systematic assessors and the unit/evidence gatherer assessors.

At the beginning of the year there was a grudging acceptance of the new curriculum documents, although principals and teachers were concerned about the number of objectives to be covered, the broadness of the objectives and the

levelling system. There was far more uncertainty associated with the Ministry's agenda in assessment. Many were convinced that all the objectives had to be assessed; that the objectives were too broad to be assessed effectively, and were unsure how to go about identifying and clarifying the knowledge and skills that would measure whether a particular objective had been achieved. They worried that the levels were going to be used for labelling students and schools.

There is no doubt that the curriculum changes and the developments in assessment have made teachers more accountable. The new national curriculum is prescriptive and the National Education Guidelines ensure that schools are held directly accountable for its delivery: it must be balanced, delivered in ways appropriate for individual students and individual progress must be monitored effectively and cumulatively. Despite the increase in accountability, schools have maintained a level of autonomy that still allows them considerable freedom, although to begin with, this was not perceived to be the case, especially with assessment.

Over the year, as the principals, the other members of the lead teams and the teachers debated the curriculum, understood more about assessment, made the connections between the curriculum and its assessment and made decisions about how they would put their new learnings into practice, they began to recapture the process for themselves. This happened in a number of ways. They were designing and trialing their own systems that they had put together in collaboration with their colleagues. They were making decisions about which objectives they would emphasise in their planning and delivery and which they would assess. They still controlled the contexts for study and, for some teachers, this was more important than deciding on the objectives. They were able to make decisions about how the curriculum was to be organised in their school and how learning areas and objectives would be integrated. They were making the curriculum work for them and the other stakeholders in their school community.

Assessment literacy

Stiggins (1991) would not describe these teachers as assessment literates. In his view, assessment literates have built-in alarm systems that sound when an assessment target is unclear, when an assessment method misses the target, when a sample of performance is inadequate, when extraneous factors are creeping into the data, and when the results are simply not meaningful to them (ibid).

At the end of the year, two respected practitioners from each of the schools were interviewed. At those interviews, the teachers were asked about the developments that had taken place across the school as well as their own individual practices. As they talked, it became clear that while they had learned a great deal over the course of the year, there were still enormous gaps in their assessment knowledge. They had gained much better understanding of the underpinning theories associated with assessment and had a much clearer grasp of its purposes and the techniques that were available to them. It became evident, however, that when they came to design assessment tasks and analyse and use the results, the teachers had little confidence in their ability to make good judgements. They were unsure about the worthiness of either the tasks or the data that was generated. What they lacked, in particular, was technical assessment expertise.

It would be unfair to lay the blame at the feet of the teachers. During the year, many teachers commented that the Contract was the first formal opportunity to develop their assessment skills since their days at Teachers' College. Some remembered their assessment training as being very technical and quantitative. One teacher described the lectures as *dry and crusty*. For others, their memories were vague. A discussion at one school, with five beginning teachers representing two pre-service institutions, revealed that they appeared to have received much broader and more comprehensive training but they concurred that they needed the support of more experienced colleagues

(experienced in assessment) to help them transfer that learning into their classrooms.

Stiggins (1991) draws a distinction between what he sees as those who need to be 'functionally literate' (students, parents and trustees) who are *users of data*, the 'practically literate' (principals and teachers) who are *generators and users*, and the 'advanced literates' (specialists in measurement and evaluation) who *generate data for the use of others*. To be 'practically literate', teachers should, at least, be able to describe the knowledge to be learned, the forms of thinking to be mastered, the particular behaviours to be demonstrated and the products to be created. Stiggins fails to include the specialist skills and techniques required by senior staff, in particular, to be able to aggregate and deaggregate assessment data, both quantitative and qualitative, for the purposes of evaluating programmes and measuring the achievement levels in specific areas of a cohort, a certain group of students or the whole school. The legislated right of Boards of Trustees to have access to that data is adding further pressure on schools to become sufficiently assessment literate to be able to provide clear and intelligent responses. The school that developed the benchmark portfolio in transactional writing could see the potential for generating valid data for this purpose.

It has long been accepted practice for New Zealand primary schools to focus on improving the quality of the programmes (undertaking professional or school-wide development) or carrying out required activities (taking the students on a camp or completing certain Maths units) rather than desired outcomes. Increasingly, the Ministry, educators and the public are demanding that educational outcomes are measured and that data be provided on the levels of achievement of our students¹⁹. What many principals and teachers have firmly rejected are calls from the government for a series of mandated standardised tests that will be used to protect standards even if, in their view,

¹⁹ The AIMHI (Achievement in Multi-Cultural High Schools) Project is just one example of a longitudinal programme where the community, schools and the government are attempting to work together to improve levels of achievement for a particular group of students.

teachers cannot do so. Unless schools can demonstrate that they have the ability to measure achievement outcomes, the process will be captured by a pencil-and-paper, large-scale and norm-referenced testing programme. Such a move would be reminiscent of the Proficiency Examination that regulated the primary curriculum from late last century through to the end of the 1930's.

The experience in England (Gipps et al, 1995) with such a formal, high-stakes programme would also suggest that other alternatives would be preferable. Teachers experienced feelings of guilt and anger when the results were published and principals and teachers were prepared to avoid this at all costs in the future; there were feelings of dissonance and alienation because of the perceived invalidity of the testing programme; and they were concerned about the impact of the tests on the students. Nonetheless, it is not unreasonable to expect that such measures be used in the absence of practitioners providing other viable and valid alternatives.

While the current crop of centrally funded ABLE Contracts is providing a forum for the improvement of assessment practices, the current study demonstrates that one year in a staff development assessment programme is insufficient to provide the teachers with all the development required to make them literate assessors. It is also important to remember that schools self-selected for the programmes and not all of those which applied were necessarily accepted. This year (1998), no contracts in assessment have been advertised and it appears that the focus has shifted from one of school-based development to national assessment. The only ongoing development work in this area that will directly benefit individual teacher practice will be the professional development associated with the National Education Monitoring Programme²⁰ (NEMP). Each year teachers are invited to participate either as task developers, administrators or markers and analysts: a small window of opportunity each year for a small group of teachers.

²⁰ NEMP is an ongoing programme of national assessment. It monitors the achievement levels and identifies the trends in students' learning in all the learning areas over a rotating, four year cycle. The monitoring involves random samples of students in years 4 and 6.

Assessment partnerships

This section examines the issue of control of the assessment process, particularly in relation to teachers and parents and teachers and students. It discusses how the schools went about redressing the balance of power so that the locus of control was not so firmly with the professionals and was shared more with the ultimate consumers of the data.

Evidence from the study supports the notion that, for a range of reasons, teachers act as gatekeepers of assessment information. Sometimes, the gatekeeping is motivated by a lack of confidence, in either pedagogical knowledge or knowledge of assessment theory and practice. At other times, it is driven by a desire to protect parents from the truth. Some teachers in the study schools expected parents not to understand what they perceived to be complex professional knowledge. Sometimes they were afraid that information would be misinterpreted or that it would be used by parents to label or stereotype their children. On that basis, they made decisions about what they considered to be 'good' or 'bad' for parents to know and the teacher, in this instance, turned benevolent guardian.

The language used in reporting to parents was raised in a survey carried out by one of the schools. The school has a mid-decile ranking and serves a relatively diverse socio-economic area. In their responses, the parents' plea, quite clearly, was for honesty. Their sentiments were echoed in an article in the New Zealand Listener where Tolerton (1997) described the vogue in the 1990's for positive reinforcement and polite circumlocution. Many parents are familiar with the percentages, place in class and medians of twenty or thirty years ago and, as indicated in the survey, some are frustrated by having to 'read between the lines' and not getting the information they need. Initial indications suggest that as teachers become more clear about what they should be reporting on (indicated by the priorities for assessment in their profiles), share the individual records they keep with parents and can negotiate and describe what can be done to progress a particular student's learning (in

other words the reporting becomes both summative **and** formative), parents appear to experience increased satisfaction with the reporting process.

While the truth of the old maxim can be argued - that parents know their children best but do not always know what is best for their children - gatekeeping of assessment information denies parents their right to make choices about how to support their child's learning. It is disempowering. In order for assessment to be a partnership between teachers and parents, five considerations emerged as critical. Firstly, teachers must have access to information in which they have confidence and which says something meaningful about what is important to learning and achievement. The second consideration is a prevailing attitude that a student's learning will be enhanced by the participation of the parents in the assessment process and that they have an undeniable right to access the information that results from that process. Thirdly, there needs to be a preparedness to be honest. While this means giving accurate information, it also involves having the skills to be able to deliver difficult messages. Teachers are trained to work with students and do not have the opportunity to develop higher order communication skills that support them to work successfully with adults. Having the confidence to communicate successfully with parents is crucial to developing an effective assessment partnership with them. The fourth consideration is enabling parents to be functionally literate with respect to data. Like others that Stiggins (1991) describes as functionally literate (students and trustees), parents need to know about the basic types of assessment, they must understand the decisions they make based on assessment data, how those decisions relate to teaching and learning, and what kinds of data can inform their decision making. They need to be sufficiently informed to be able to ask the questions necessary to promote sound assessment. The final consideration is designing and timetabling a reporting system that takes cognisance of parents. For example, the school that asked the questions *What do parents need to know?* and *When do they need to know it?* made some radical changes to practices long steeped in tradition and what had suited the teachers rather than the parents. These changes included reporting more often, making use of original documentation

rather than preparing specific parent reports and engaging parents more directly in action that resulted from the reporting process. There was also evidence to suggest higher levels of parental and teacher satisfaction with the process and indications of higher parental regard for teachers as professionals as a result of focussing more on meeting parents' needs.

The discussion now turns to the assessment partnership between students and teachers. To begin with, some teachers viewed involving students in new ways, as building in an additional activity into their planning to add variety. It quickly became evident that if teachers were going to treat these initiatives seriously they needed professional development and support to understand the importance of meta-cognitive skills and the crucial significance of self-review and ownership of the learning process to student achievement. It could not be assumed that all teachers would intuitively have these understandings and skills upon which to draw. For example, some teachers were surprised to learn about the developmental progressions in goal setting and gained considerable confidence in sharing strategies to guide students through the process. They needed information. Others were fearful of including students in the teacher/parent conferences and were concerned that would not have the skills to manage what, for some of them, were already tense situations. These teachers needed skills. Many teachers had learned how to manage conferences through experience and had never reviewed their own performance in that situation or had the opportunity to discuss in-depth with their colleagues how they went about their conferences. Some of this group were resistant to making any changes in this regard. These teachers needed to see alternative models to gain insights into other ways of operating. The surprise for the lead teams was the number of issues this aspect of assessment generated for discussion and professional development. Assessment partnerships with students will not happen automatically or on suggestion.

Assessment partnerships also need to encompass involvement with other schools. While the new curriculum aspires to provide a seamless education system (Ministry of Education, 1993b), there is evidence, at least in this study,

that some teachers pay lip service to assessment data provided by other teachers. Teachers do not always trust the information that is handed on to them from other schools and can be very critical of the data the records contain. The stories of secondary school vaults or storage areas under stages bulging with unread or barely read records from intermediate schools are legend amongst the teaching fraternity. When assessment data is not used, the disjuncture that results has the potential to create a barrier to achievement, even though it may only be temporary, and to put a student's learning at risk. As described in the previous chapter, all four of the study schools worked on aspects of assessment that included developing school-wide systems. Three of them extended their current recording system or designed new systems, using either profiles or portfolios. All were tailor-made for each school. While there were some similarities in either the type of format or the type of data teachers were collecting, particularly with the profiling records, making comparisons between students from the different schools was not easy.

This raises some important issues for students at key transition points (at the end of year 6 and year 8) and when students move to a new school at any other point in their schooling. While the new records being developed by the schools were giving them greater consistency *within* their schools, problems *between* the schools appeared to be compounding. The disparity in formats and the length of some of the documents were deterrents to teachers using them. Similarly, the lack of quantitative data, especially in the portfolios, meant that senior staff felt that snap judgements, for placement purposes, could not be readily made. Consistency of basic information on transition remains an unresolved issue. At this point, each school is operating as an independent 'educational island' and finding ways to build bridges between them, at least between schools that serve the same local area, makes pedagogical sense. The sharing of assessment information is one issue for the agenda.

Methodological insights

As stated by Poskitt (1994), action research in New Zealand is in its infancy, with few models to emulate and professional development only recently incorporating action research principles. This section will discuss aspects of the study that provided some insights into the action research process as it applies in New Zealand schools, outline some techniques and structures that appear to support practitioner ownership of the action research process and elaborate on some theoretical positions taken by some academics in relation to first and second order domains of action research.

Action research was a new concept to almost all the principals and the teachers in the four study schools. There was not an embedded research culture in any of the schools and even reading a short research article was a new experience for many of them. They associated research with the academic world rather than that of the practitioner and saw it as shrouded in esoteric language and abstract concepts that have little to do with the day-to-day realities of the classroom. They did not have the view that they were capable of 'doing' research even though they related to the model and could see the similarities between it and models of school development with which they were familiar. One of the most useful techniques in supporting the development of a research culture and, at the same time, serving as a powerful way of the schools taking responsibility for the research (as first-order researchers), was the Help Desk. It was easy for them to take the initiative to access the researcher, get the resources, information or support they needed and then take responsibility for acting on them.

One of the greatest barriers to practitioner research for these teachers was their lack of any knowledge of basic research data collection techniques. The irony was that, if they had had training and support in this area in assessment, then the transfer of those skills to research would have made this less of an issue. Johnston (1994) cites this as a reason for abandoning action research in favour of narrative inquiry. Story-telling was an important component of the

action research process in which the teachers engaged but, in this instance, some of the most telling and profound learnings of the teachers occurred as a result of the data collection and its analysis (eg the baseline questionnaire, the survey of parents' perceptions of one school's reporting system; the analysis of teachers' individual student profiles to define the quality of the qualitative data being recorded). What the research highlighted was that teachers are not familiar with a range of data gathering mechanisms nor do they have the skills to collate and analyse that data, to either support their students' learning or their own. Zuber-Skerritt (1993: 55) warns against seeing the techniques as an end in themselves but suggests they are a prerequisite for developing higher levels of inquiry. There is an urgent need for New Zealand teachers to acquire these skills.

Another aspect of action research that was foreign to some of the teachers was critical reflection, even at a group level. The workshops were established with the notion of developing a dialogical community (McNiff, 1988). Some of the most important discussions, vigorous debates and significant mind-shifts occurred in these meetings. There were several aspects to the groups that made this level of reflection possible.

Firstly, they had a common interest: assessment. Secondly, the group established its purpose and the ground rules for operation at the very first meeting. They were set up with a clear and express purpose of being forums for critique and reflection. Thirdly, the meetings were held in a quiet venue away from the schools where the participants had uninterrupted thinking time. Finally, each lead team brought an objectivity to the discussion of each other's practices that could not be achieved to the same extent back in the individual schools. The latter was critical to the success of the school-based work. Lead team members from other schools could, and did, ask questions and raise issues that no one back at their own school would think or dare to ask. They became second-order researchers in that process and learned to hone some of the research skills they needed to apply to their own processes. Here, samples, documents and data were put up for discussion for which each Lead

Team took collective responsibility. In contrast, it was interesting to observe the nervousness of the teachers in disclosing their individual practices in public ways back at the individual schools. While they were happy to have opportunities to **talk** about their practice, they were more reserved about **demonstrating** that practice by showing samples of work, sharing comments written in students' books and looking at each other's reports or profiles. Yet this was an opportunity to take change to yet another level. It suggests that objectivity, however it is achieved, is critical in the early stages of practitioner research.

For these teachers, one of the critical aspects of the process was its cyclic and systematic approach. This is not to suggest that any of the schools' processes were tidy and smoothly executed. They were the opposite: often untidy, sometimes muddled and frequently interrupted. Johnston (*ibid*) argues that the systematic nature of action research is, in some ways, limiting and 'unnatural'. This was not the experience of these schools. The 'model' was used in discussions by the teachers as a touchstone for maintaining momentum, for guiding them forward and keeping them on track. It allowed them to be responsive to the context while at the same time providing them with a framework for collective action.

Claims that action research needs or promotes collaboration are easy to come by (Hustler et al, 1986; McNiff, 1988; Elliot, 1991), although others (Johnson, 1994; Waters-Adams, 1994) argue that the political nature of staffrooms can make collaboration difficult, if not impossible, to achieve. There is no doubt, that the personal dynamics within the four schools slowed down the process at times, occasionally took the process in a different direction from that intended and, in one instance stopped any progress being made for close to a term. Without the support of the outside researcher, some of these issues would have been difficult to resolve or by-pass. However, what all schools achieved was the introduction of at least one school-wide innovation that, in three instances, involved a major departure from the previous practice of most of the teachers. To achieve the change involved cooperation. To achieve the depth

of change that occurred in some of the schools (the process of constructing the benchmark portfolios, instituting the new reporting system, improving the quality of feedback to students in the portfolios) required collaboration. It was essential to making the changes as quickly as, and to the extent, they did.

The discussion now turns to the tension often described in the literature between the first and second-order domains of action research. Some imply that an outside researcher in some way distorts the research and reduces the emancipatory potential of the process (Carr and Kemmis, 1986). Poskitt states (1994) that prior to the first-order domain being developed, it seems that schools of limited research experience may benefit from involvement in second-order research. This suggests a 'sliding scale' of second-order research leading to the practitioners gradually taking over the process as first-order researchers. This can occur as they learn more about research, gain the skills needed and gather confidence. Suggesting the desirability of little or no 'interference' from outside or that the first-order domain ceases or fades as the practitioners take full responsibility for the process, precludes a needs-based view of how the action research process might take place. Varying degrees of outsider help may be needed from time to time, regardless of the level of skill of the group, and depending on the complexity of the issue with which the group is grappling or the dynamics of the people involved at the time. Reification of any particular model or process suggests adherence to 'a certain way of knowing' that contradicts the very spirit of naturalism underpinning action research.

Finally, it is important to refer back to the research questions that provided the original framework for the study. As was stated in the introductory chapter, the questions were formulated to act as a guide to the research rather than to constrain or control it in any way and at no time were they ever shared with the schools. Great care was taken to ensure that the schools controlled the research process as much as possible, that the aspects of assessment they selected to work on were identified by them and that the project centred on ensuring that their needs were met. With these criteria in mind, the questions

were designed to be sufficiently broad as to accommodate the problems and issues identified by the schools. The most important functions of the questions were to help shape the researcher's thinking prior to the work beginning in the schools and as a reference point when 'mapping' the data into themes. Nonetheless, the themes that emerged are a direct reflection of the particular choices the schools made and the processes they followed rather than the research questions. All of the questions have been addressed to a greater or lesser extent in this and previous chapters but the answers are woven into a research journey that mirrors the work of the schools rather than the researcher's attempts to anticipate how that journey might unfold.

Future research

Given the dearth of research into the classroom assessment practices of New Zealand teachers and the specific issues raised in this study, there is ample scope for research into a range of aspects of assessment, both formative and summative. These could include the giving of oral and written feedback to students; the impact on motivation of formative feedback, self and peer assessment; tracking developments in the assessment literacy of the teachers including the aggregation of data for review and reporting purposes; and investigating further efforts by teachers to engage in moderation as evidenced in the study with the developments in benchmarking.

As has been illustrated by this research, the recording methods used by teachers over the years have been a gauge of the philosophies of and approaches to assessment that have dominated the thinking of governments and teachers at the time. Currently, schools have captured these recording processes and are developing individual systems that reflect their own particular views of assessment and which, they believe, best meet the needs of the stakeholders in their schools. It will be important to monitor how these systems change and develop over time and what influences those changes. The willingness to share information that results from these assessment

processes and to involve parents more fully in the assessment process could also be a significant area of study.

It could also be important to monitor the impact of the new curricula, introduced since the study concluded, in Technology, Social Studies, Physical Health and Well-Being and the Arts. Traditionally, these areas have been more problematic for teachers in relation to assessment because they have required more specialist knowledge (eg Art and Music) or because the outcomes are perceived to be more difficult to measure (eg Social Studies). Teachers have traditionally used the framework of learning areas or subjects for their assessment. With the growing emphasis on meta-cognition, which motivated one school in this study to emphasise portfolios and self-assessment, it could also be significant to monitor whether schools continue to measure student progress and achievement on the basis of 'subjects' or whether there will be a shift towards framing assessment around the essential skills as outlined in the Curriculum Framework (Ministry of Education 1993a).

Increasingly, there are demands for schools and teachers to be made more accountable, not only for curriculum delivery but for achievement outcomes of students. A level of tension exists between giving schools autonomy and the opportunity to be self-managing and the accountability of those schools to a government driven by a New Right ideology. This ideology suggests that only those that 'perform' to expected standards should survive. Will the new national curriculum and the meeting of legislative requirements of Boards of Trustees, particularly in relation to assessment and student achievement, be sufficient to satisfy the demands for public accountability? The suggestion that a system of national testing be introduced (Ministry of Education, 1998, 1) suggests that, for this government at least, it is not sufficient.

Summary

Students of organizational culture recognise that resistance to innovation is deeply rooted in individual psychology and group culture (Schein, 1985). We

exalt it in principle but oppose it in practice. Evans (1996) suggests that reforming practice inevitably involves a double standard: when we advocate change, we usually mean by other people. This research demonstrates that when academics and practitioners work together in a participatory and democratic way, quite profound learning takes place and radical changes can occur that alter the practices of both groups. It also signals that there is an enormous amount of work to do if New Zealand students are to benefit from the assessment research that is shaping learning and teaching both in this country and around the globe. This thesis highlights some of those challenges for principals, teachers and researchers as we continue to rethink assessment.

BIBLIOGRAPHY

- Altrichter, H., Posch, P. and Somekh, B. (1993) *Teachers Investigate their Work: An introduction to the methods of action research*. London: Routledge.
- Bailey, J. and McTigue, J. (1996) Reporting achievement at the secondary level: What and how, in the *ASCD Yearbook*. Alexandria, Virginia: ASCD.
- Bennett, R., Gottesman, D. and Cerullo, F. (1993) Influence of behaviour perceptions and gender on teachers' judgements of students' academic skill. *Journal of Educational Psychology*, 85, 347-356.
- Bishop, J.H. (1992) Why students need incentives to learn. *Educational Leadership*, 49 (6), 15-18.
- Black, P.J. (1993) Formative and summative assessment by teachers. *Studies in Science Education*, 21, 49-97
- Black, P. and Wiliam, D. (1998) Assessment and classroom learning. *Assessment in Education*, 5 (1), 7-74.
- Blenkin, G., Edwards G. and Kelly, A. (1992) *Change and the Curriculum*. London: Paul Chapman Publishing Ltd.
- Blackmore, J. (1988) *Assessment and Accountability*. Geelong, Victoria: Deakin University Press.
- Brewer, W.R. and Kallick, B. (1996) Technology's promise for reporting student learning, in the *ASCD Yearbook*. Alexandria, Virginia: ASCD.
- Broadfoot, P. (1992) *Changing priorities in assessment*. Paper presented at the Conference on Qualifications for the 21st Century, Wellington.

- Brookhart, S. (1993) Teachers' grading practices: Meaning and values. *Journal of Educational Measurement*, 30(2), 123-142.
- Brown, A., Campione, J., Webber, L. and McGilly, K. (1992) Interactive Learning Environments: A new look at assessment and instruction, in Gifford, B. and O'Connor, M. (eds) *Changing Assessments: Alternative Views of Aptitudes, Achievement and Instruction*. Boston: Kluwer Academic.
- Bruner, J. (1985) Vygotsky: an historical and conceptual perspective, in Wertsch, J. (ed) *Culture, Communication and Cognition: Vygotskyian perspectives*. Cambridge: Cambridge University Press.
- Burgess, R. (ed) (1989) *The Ethics of Educational Research*. New York: The Falmer Press.
- Burke, K. (1992) *Authentic Assessment*. Australia: Hawker Brownlow Education.
- Carr, W. and Kemmis, S. (1986) *Becoming critical: Knowledge and action research*. London: Falmer Press.
- Costa, A., and Kallick, B. (eds) (1995) *Assessment and the learning organisation: Shifting the paradigm*. Virginia: ASCD Press.
- Croft, C. (1995) Assessment strategies for the New Zealand Curriculum Framework. *Set 1, Item 13*.
- Crooks, T. (1988) The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438-481.
- Crooks, T. (1996) Threats to the valid use of assessments. *Assessment in Education*, 3(3), 265-285.

Dalin, P. and Rolff, H. (1993) *Changing the School Culture*. London: Cassell.

Darling-Hammond, L. (1994) National standards and assessment: Will they improve education? *American Journal of Education*, 102, 478-509.

Department of Education (1987) *The Curriculum Review*. Wellington, Government Printer.

Department of Education. (1989a) *Assessment for Better Learning*. Wellington, Government Printer.

Department of Education (1989b) *Keeping School Records*. Wellington, Government Printer.

Diez, M. and Moon, C. (1992, May) What do we want students to know?...and other important questions. *Educational Leadership*, 49(8), 38-41.

Education Review Office. (1995) *Assessing Student Achievement*. Autumn, 3.

Education Review Office (1996) *Reporting Student Achievement*. Autumn, 3.

Eisner, E. (1993) Reshaping assessment in education: some criteria in search of practice. *Journal of Curriculum Studies*, 25(3), 219-233.

Elam, S., Rose, L. and Gallup, A. (1991) The 23rd annual Gallup Poll of the public's attitudes towards public schools. *Phi Delta Kappan*, 73 (1): 41-53.

Elam, S., Rose, L. and Gallup, A. (1992) The 24th annual Gallup/Phi Delta Kappa Poll of the public's attitudes towards public schools. *Phi Delta Kappan*, 74 (1): 41-53.

Elam, S., Rose, L. and Gallup, A. (1994) The 26th annual Gallup/Phi Delta Kappa Poll of the public's attitudes towards public schools. *Phi Delta Kappan*, 76 (1): 41-53.

Elley, W. (1995) What is wrong with standards-based assessment?, in Peddie, R. and Tuck, B. (eds) *Setting the Standards: Issues in Assessment for National Qualifications*. Palmerston North: The Dunmore Press.

Elliot, J. (1987) Educational theory, practical philosophy and action research. *British Journal of Educational Studies*, 35 (2), 149-169.

Elliot, J. (1988) Teachers as Researchers: Implications for Supervision and Teacher Education. American Educational Research Association address, New Orleans.

Elliot, J. (1988) Educational research and outsider-insider relations. *Qualitative Studies in Education*, 1(2), 155-166.

Elliot, J. (1991) *Action Research for Educational Change*. Milton Keynes: Open University Press.

Elliot, J. (1993) *What is good action research? - some criteria*. Paper presented at the Fourth Annual Action Research Symposium. Puget Sound Educational Consortium, University of Washington, Seattle. April, 1993.

Evans, R. (1996) *The Human Side of School Change*. San Francisco, California: Jossey Bass.

Gardner, H. (1992) Assessment in context: The alternative to standardised testing, in Gifford, B. and O'Connor, M. (1996) *Changing assessments : Alternative Views of Aptitude, Achievement and Instruction*. Boston: Kluwer Academic Publishers.

Geertz, C. (1973) *The Interpretation of Cultures*. New York: Basic Books.

Gifford, B., and O'Connor, M. (1996) *Changing Assessments: Alternative Views of Aptitude, Achievement and Instruction*. Boston: Kluwer Academic Publishers.

Gipps, C. (1994) *Beyond Testing: Towards a theory of educational testing*. London: The Falmer Press.

Gipps, C., Brown, M., McCullum, B. and McAllister, S. (1995) *Intuition or Evidence*. Open University Press.

Gipps, C., McCullum, B. and Brown, M. (1996) Models of teacher assessment among primary teachers in England. *The Curriculum Journal*, 7(2), 7-183.

Glassman, N. and Glassman, L. (1988) Educational reform and evaluation. *Education Administration Quarterly*, 24(4), 438-445.

Guskey, T. (ed) (1996) Communicating student learning, in *ASCD Yearbook*. Alexandria, Virginia: ASCD Press.

Hargreaves, A. (ed) (1997) *Rethinking Educational Change with Heart and Mind*. Virginia: ASCD Press.

Harlen, W. (1991) National Curriculum assessment: Increasing the benefit by reducing the burden. *Education and Change in the 1990s*, Journal of the Educational Research Network of Northern Ireland, 5, February, 3-19.

Harlen, W., Gipps, C., Broadfoot, P., and Nuttal, D. (1992) Assessment and the improvement of education. *The Curriculum Journal*, 3(3), 215-230.

Hebert, E. and Schultz, L. (1996) The power of portfolios. *Educational Leadership*, 53(7), 70-71.

Herman, J. (1992) What good research tells us about good assessment. *Educational Leadership*, May, 74-78.

Hill, M. (1997, December) *Models of practices of teachers in New Zealand primary schools*. Paper presented at the annual conference of the New Zealand Association for Educational Research, Auckland.

Hossack, L.A. (1997) An action research primer for principals. *The Canadian School Executive*. January, 9-13.

Hustler, D., Cassidy, A. and Cuff, E. (eds) (1986) *Action Research in Classrooms*. London: Allen and Unwin.

Irving, J. (1992) *Educational Assessment and Monitoring in New Zealand*. Wellington: Ministry of Education.

Irving, J. (1993, September) *Policies and future directions for assessment of the individual and the system in New Zealand education*. Paper presented at the Conference Assessment and Learning in New Zealand: Challenges and Choices, Palmerston North College of Education, Palmerston North.

Irving, J. (1997) *School-wide assessment: The big picture*. Wellington: New Zealand Council for Educational Research.

Johnston, S. (1993) Action research as a school-level change process. *Curriculum Perspectives*, 13(1), 21-28.

Johnston, S. (1994) Is action research a 'natural' process for teachers? *Educational Action Research*, 2(1), 39-48.

Kemmis, S. and McTaggart, R. (eds) (1988) *The Action Research Planner (3rd ed)*. Geelong, Victoria: Deakin University Press.

Khatti, N., Kane, M. and Reeve, A. (1995) How performance assessments affect teaching and learning. *Educational Leadership*, November: 80-83.

Kulik, C., Kulik, J. and Bangert-Drowns, R. (1990) Effectiveness of mastery learning programmes: A meta-analysis. *Review of Educational Research*, 60(2), 265-299.

Kyle, D., and McCutcheon, G. (1984) Collaborative research: Development and issues. *Journal of Curriculum Studies*, 16(2), 173-179.

Lawton, D. (ed) (1989) *The Education: Choice and Control*. London: Hodder and Stoughton.

Le Countryman, L. and Schroeder, M. (1996) When students lead parent-teacher conferences. *Educational Leadership*, April, 64-68.

Lewin, K. (1948) *Resolving Social Conflicts*. New York: Harper.

Little, A. and Allan, J. (1989) Student-led parent-teacher conferences. *Elementary School Guidance and Counselling*, 23: 210-218.

Lomax, I. (1991) Clarifying some issues about action research: a personal response, in Edwards, G. and Rideout, P. (Eds) (1991). *Extending the Horizons of Action Research*. University of East Anglia, Norwich, Classroom Action Research Network.

McAlpine, D., Poskitt, J. and Bourke, R. (1998) *Assessment for Better Learning: Final Report*. Wellington, New Zealand. Ministry of Education.

McCulloch, G. (ed) (1992) *The School Curriculum in New Zealand: History, Theory, Policy and Practice*. Palmerston North, New Zealand: The Dunmore Press.

McKernan, J. (1997) *Curriculum Action Research: A Handbook of Methods and Resources for the Reflective Practitioner*. London: Kogan Page.

McNiff, J. (1988) *Action Research: principles and practice*. Basingstoke: McMillan.

McTaggart, R. (1991) *Reflection on teaching: Creating an enquiry in education*. Address presented as Lansdowne Visitor at the University of Victoria, British Columbia, Canada. March 25, 1991.

Madaus, G. (1988) The influence of testing on the curriculum. *Critical Issues in Curriculum*, 87th Yearbook of NSSE, Part 1. Chicago: University of Chicago Press.

Meyer, C. (1992) What's the difference between authentic and performance assessment. *Educational Leadership*, May.

Ministry of Education. (1992) *The National Curriculum of New Zealand: a discussion document*. Wellington: Ministry of Education Learning Media.

Ministry of Education. (1993a) *The New Zealand Curriculum Framework*. Wellington: Ministry of Education Learning Media.

Ministry of Education. (1993b) The National Education Guidelines. *The Education Gazette*. 72(7), 3.

Ministry of Education. (1994) *Assessment: Policy to Practice*. Wellington: Ministry of Education.

Ministry of Education. (1995) Assessment issues. *The Education Gazette*. 74(5), 1-2.

Ministry of Education. (1998) National assessment policies and programmes. *The Education Gazette*. 77(19), 1.

Moss, P. (1992) Shifting concepts of validity in educational measurement: Implications for performance assessment. *Review of Educational Research*. 62(3), 229-258.

Murphy, R. (1990) National assessment proposals: Analysing the debate, in Flude, M. and Hammer, M. (eds) *The Education Reform Act 1988*. London: The Falmer Press.

Nias, J. (1993) Changing times, changing identities: Grieving for a Lost Self in Burgess, R. (ed) *Educational Research and Evaluation for Policy and Practice?* London: the Falmer Press.

Nuttall, D. (ed) (1986) *Assessing Educational Achievement*. Lewes: The Falmer Press.

O'Neill, A (ed) (1996) Curriculum development in Aotearoa New Zealand: An editorial introduction. *Delta: Policy and Practice in Education*, 48 (1), 3-10.

O'Neil, J. (1993a) The promise of portfolios. *ASCD Update*, 35(7): 1.

O'Neil, J. (1993b) Portfolio assessment bears the burden of popularity. *ASCD Update*, 35 (8):1.

Oja, S. and Smulyan, L. (1989) *Collaborative Action Research: A Developmental Approach*. London: Falmer Press.

Openshaw, R., Lee, G. and Lee, H. (1993) *Challenging the Myths: Rethinking New Zealand's Educational History*. Palmerston North, New Zealand: The Dunmore Press.

Piper, K. (1997) *Curriculum Reform and the National Interest*. Australian Council for Educational Research.

Poskitt, J. (1994) *Research as Learning: The Realities of Action Research in a New Zealand Individualised Learning Programme*. An unpublished Ph.D. thesis, Massey University, Palmerston North.

Ramsey, P., Hawk, K., Harold, B., Marriott, R. and Poskitt, J. (1993) *Developing Partnerships: Collaboration Between Teachers and Parents*. Wellington: Ministry of Education Learning Media.

Redding, N. (1992) Assessing the big outcomes. *Educational Leadership*, May.

Renwick, M., and Gray, A. (1995) *Implementing the New Zealand Curriculum in Primary Schools*. Wellington: New Zealand Council for Educational Research.

Renwick W. L. and Ingham, L. J. (eds) (1974) *Educational Planning in New Zealand*. Wellington: Government Printer.

Resnick, L. (1987) *Education and Learning to Think*. Washington, DC: National Academy Press.

Resnick, L.B. and Resnick, D.P. (1992) Assessing the thinking curriculum: New tools for education reform, in Gifford, B. and O'Connor, M. (eds) *Changing assessments: Alternative views of aptitude, achievement and instruction*. London: Kluwer Academic Publishers.

Ritchie, G., and Carr, K. (1992) A constructivist critique of mastery learning in Mathematics. *New Zealand Journal of Educational Studies*, 27(2), 191-201.

Rowntree, D. (1989) *Assessing Students: How shall we know them?* New York: Kogan Page.

Sadler, R. (1989) Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-44.

Sadler, R. (1992) *Standards-based Assessment in the Secondary School: The Queensland experience*. Paper presented at the conference 'Qualifications for the 21st Century', January, Wellington, New Zealand Qualifications Authority.

Schein, E. (1985) *Organisational Culture and Leadership*. San Francisco: Jossey-Bass.

Snook, I. (1994, March) *Challenges to Schooling*. Paper presented to the PPTA Regional Conference of Branch Managers: Auckland Region, Auckland, New Zealand.

Stenhouse, L. (1993) *Authority, Education and Emancipation*. London: Heinemann.

Stiggins, R. (1991) Assessment Literacy. *Phi Delta Kappan*, 72(7), 534-539.

Stiggins, R.J. (1994) *Student-Centred Classroom Assessment*. New York: Macmillan.

Sullivan, K. (1994) The impact of educational reform on teachers' professional ideologies. *New Zealand Journal of Educational Studies*, 29(1), 3-19.

Sutton, R. (1994) *Assessment: A Framework for Teachers*. London: Routledge.

Sutton, R. (1995) *Assessment for Learning*. Salford, England: RS Publications.

- Tolerton, J. (1997) F for Effort. *The New Zealand Listener*, December 20, 36.
- Torrance, H. (1993) Formative assessment: Some theoretical problems and empirical questions. *Cambridge Journal of Education*, 23(3), 333-343.
- Towler, L., and Broadfoot, P. (1992) Self-assessment in the primary school. *Educational Review*, 44(2), 137-151.
- Tunstall, P. and Gipps, C. (1995, June) *Teacher Feedback to Young Children in Formative Assessment: A Typology*. Paper presented to the IAEA Conference, Montreal.
- Waters-Adams, S. (1994) Collaboration and action research: A cautionary tale. *Educational Action Research*, 2(2), 195-210.
- Watt, M., and Watt, D. (1993) Teacher research, action research: The Logo action research collaborative. *Educational Action Research*, 1(1), 35-63.
- White, R. (1992) Implications of recent research on learning for curriculum and assessment. *Journal of Curriculum Studies*, 24(2), 153-164.
- Wiggins, G. (1993) *Assessing Student Performance*. San Francisco: Jossey-Bass.
- Wiggins, D. (1995) Honesty and fairness: Towards better grading and reporting. *ASCD Yearbook: Communicating Student Learning*. Virginia, USA: ASCD.
- Willis, D. (1992) Educational assessment and accountability: A New Zealand case-study. *Journal of Educational Policy*, 7(2), 205-221.

Willis, D. (1994) School-based assessment: Underlying ideologies and their implications for teachers and learners. *New Zealand Journal of Educational Studies*, 29(2), 161-174.

Willis, S. (1996) On the cutting edge of assessment: Testing what students can do with knowledge. *Education Update*, 38(4), 1-7.

Willis, D. and Bourke, R. (1998) *Catching the Slow Train: Teacher development programmes and real change*. Paper presented at the 20th Annual NZARE Conference, Otago University, Dunedin, New Zealand, December 3-6.

Winter, R. (1989) *Learning from Experience: Principles and Practice in Action Research*. London: The Falmer Press.

Wolf, D. (1987/1998) Opening up assessment. *Educational Leadership*, 44, 24-29.

Wylie, C. (1997) *Self-Managing Schools Seven Years On: What have we Learnt?* Wellington: New Zealand Council for Educational Research.

Zeichner, K. and Noffke, S. (1998) Practitioner research, in Richardson, V. (ed) (1998) *Handbook on Research in Teaching*. Washington, D.C.: American Educational Research Association.

Zuber-Skerritt, O. (1993) Improving learning and teaching through action learning and action research. *Higher Education Research and Development*, 12(1), 45-58.

APPENDIX A

National Administration Guidelines

NATIONAL EDUCATION GUIDELINES

NATIONAL ADMINISTRATION GUIDELINES (NAGS)

In order to ensure that the National Education Goals are met, Board of Trustees and Principals respectively, are also required to follow sound governance and management practices involving curriculum, employment, financial and property matters applying to schools. Further details of these requirements are found in the relevant legislation, appropriate contracts of employment and, from time to time, guidelines promulgated by the Secretary of Education.

1. Boards of Trustees must foster student achievement by providing a balanced curriculum in accordance with the national curriculum statements* (i.e. the New Zealand Curriculum Framework and other documents based on it). In order to provide a balance programme, each Board, through the Principal and staff, will be required to:
 - i. implement learning programmes based upon the underlying principles, stated essential learning areas and skills, and the national achievement objectives; and
 - ii. monitor student progress against the national achievement objectives; and
 - iii. analyse barriers to learning and achievement; and
 - iv. develop and implement strategies which address identified learning needs in order to overcome barriers to students' learning; and
 - v. assess student achievement, maintain individual records and report on student progress; and
 - vi. provide appropriate career information and guidance for all students, with a particular emphasis on specific career guidance for those students who, nearing the end of their schooling, are at risk of becoming unemployed.

Note: * Existing syllabuses are to be regarded as national curriculum statements until they are replaced.

2. According to the legislation on employment and personnel matters, each Board of Trustees is required in particular to:
 - i. develop and implement personnel and industrial policies, within policy and procedural frameworks set by the Government from time to time, which promote high levels of staff performance, use educational resources effectively and recognise the needs of students;

- ii. be a good employer as defined in the State Sector Act 1988 and comply with the conditions contained in employment contracts applying to teaching and non-teaching staff.
3. According to legislation on financial and property matters, each Board of Trustees is also required in particular to:
 - i. allocate funds to reflect the school's priorities as stated in the Charter;
 - ii. monitor and control school expenditure, and ensure that annual accounts are prepared and audited as required by the Public Finance Act 1989 and the Education Act 1989;
 - iii. comply with the negotiated conditions of any current asset management agreement, and implement a maintenance programme to ensure that the school's buildings and facilities provide a safe, healthy learning environment for students.
4. Each Board of Trustees is also required to:
 - i. document how the National Education Guidelines are being implemented;
 - ii. maintain an ongoing programme of self-review.
5. Each Board of Trustees is also required to:
 - i. provide a safe physical and emotional environment for students;
 - ii. comply in full with any legislation currently in force or that may be developed to ensure the safety of students and employees.
6. Each Board of Trustees is also expected to comply with all general legislation concerning requirements such as attendance, the length of the school day, and the length of each school year.

APPENDIX B

**Letter of introduction to the Contract
and Expression of Interest form**



MASSEY
UNIVERSITY

ALBANY

TAKAPUNA CENTRE

1 September 1995

Dear Principal

The Educational Research and Development Centre of Massey University's Albany Campus have been contracted by the Ministry of Education to provide a school-based professional development programme in assessment during 1996. The Contract facilitators will be Jan Hill and Kay Hawk.

The programme will involve 12 Primary and/or Intermediate schools and 4 Secondary schools from the Auckland area. The Principal and two Designated Teachers from each school will work with the ERDC facilitators to coordinate and facilitate the development programme in their school.

The purpose of the contract is to:

- help each school to critically analyse their current assessment practices
- provide up-to-date information on assessment theory, issues and practices
- provide support to put in place effective and manageable assessment practices
- encourage teachers to integrate assessment practices with the learning and teaching process
- encourage the development of assessment practices that meet the requirements of the National Education Guidelines

Based on a highly successful assessment contract currently being run by ERDC, the programme will consist of :

- a one-day Induction Workshop
- a minimum of three half-day school visits where the facilitators will plan and work with the Principal, the two designated teachers and possibly other key staff or the whole staff; as needs arise, and on the school's request, more visits may be organised.

- two full-day and three half-day Development Group Workshops (schools will be divided into four Development Groups based on the type of school and/or the assessment issues they wish to work on)
- three mailouts of articles and readings on assessment
- access to a Help Desk where schools can phone or fax for additional information and support
- on-going work in each school coordinated by the Principal and two Designated Teachers

More specific details are provided in the programme outline attached.

Each school will receive **nine** TRDs to attend all the workshops and meetings and it is expected that each school will match that with funding from their own Staff Development budget to continue the development work back in their schools.

If assessment is a curriculum priority for your school and you would like to be considered for inclusion in this Contract please complete the Expression of Interest Form attached and return it no later than **Friday, September 29** to:

ERDC
Massey University
PO Box 331-443
TAKAPUNA

Kind Regards



MASSEY
UNIVERSITY

ALBANY

TAKAPUNA CENTRE

ASSESSMENT FOR BETTER LEARNING

EXPRESSION OF INTEREST

1. PRINCIPAL'S NAME _____ SCHOOL _____
2. STREET ADDRESS _____ POSTAL _____

3. PHONE NUMBER _____
4. FAX NUMBER _____

5. PLEASE INDICATE BELOW THE ASSESSMENT ISSUES YOUR SCHOOL WISHES TO WORK ON DURING 1996.

TICK THOSE THAT APPLY

- | | |
|--|---|
| <input type="checkbox"/> developing profiles to record pupils' progress | <input checked="" type="checkbox"/> developing an assessment policy |
| <input checked="" type="checkbox"/> assessing the essential skills | <input checked="" type="checkbox"/> developing more streamlined/effective recording and reporting processes |
| <input type="checkbox"/> developing assessment techniques in a specific subject area. Please state which subject : _____ | <input type="checkbox"/> identifying and addressing barriers to learning |
| <input type="checkbox"/> developing Records of Achievement | <input type="checkbox"/> Others (please specify) _____ |
| <input type="checkbox"/> student involvement in the assessment process | _____ |
| <input type="checkbox"/> aggregation of assessment information across the school/department | _____ |
| <input checked="" type="checkbox"/> increasing teachers' understanding and use of a range of assessment techniques | |

TICK ONE BOX

5. ☐ This has been discussed with Senior Staff
- ☒ This has been discussed with the whole staff
- ☐ This is my idea at this stage

6. Number of full time teachers in the school including the Principal _____

PLEASE RETURN TO: Assessment for Better Learning Contract
 ERDC, Massey University
 PO Box 331-443
 TAKAPUNA

Before Friday, September 26th, 1995

APPENDIX C

Baseline post-box question questionnaire

ASSESSMENT FOR BETTER **LEARNING CONTRACT 1996**

This feedback will contribute directly to help the school plan developments in
assessment for this year

1. My greatest concerns about student assessment are . . .

2. In my view, the key purposes assessment should serve are . . .

3. Assessment techniques I feel very knowledgeable about and
confident in using are . . .

4. Some questions about assessment I would like answered are . . .

5. Something that may make it difficult to change the way we work in assessment in this school is . . .

These questionnaires will be confidential to ERDC but a summary of the results will be given back to the school.

APPENDIX D

Sample action plans

Objective	Action	Responsibility	Time Frame	Completed
To get all teachers understanding the expectations of the contract and their role in it.	Staff meeting No. 1.	H	April 24	
To get all teachers understanding the requirements and the key consideration that under pin assessment.	P11 - 13 Assess P to P (requirements) Highlight how we are meeting MAGS? Any Gaps?	C	April 24	
	Ed Gazette article - Surprises - Interesting Points - Questions Key considerations	A	April 24	
To clarify teachers understanding about the purpose of portfolios.	OHT Crookes 22 thesis. Select up to 11 articles for sharing with staff.	H CHA	April 24 May 1	

Objective	Action	Responsibility	Time Frame	Completed
To collect feedback on the portfolios as the basis for a review.	<p>Staff Meeting No. 2.</p> <ol style="list-style-type: none"> 1. Divide articles amongst staff. 2. Read, highlight write questions. 3. Discuss & record key ideas. <ul style="list-style-type: none"> - Questions - Surprises - Interesting points 4. Summarise relevant key Ideas applicable to Kowhai 			
	<p>Staff Meeting No. 3.</p> <ol style="list-style-type: none"> 1. Divide the staff into mixed groups. <ul style="list-style-type: none"> - Students - Parents - Secondary Schools - Staff. 		May 15	
	<ol style="list-style-type: none"> 2. Groups to decide how to collect the information. <ul style="list-style-type: none"> - Size of group - Technique <ul style="list-style-type: none"> • Survey • Group interview - When - Who - Time 		May 15	

Objective	Action	Responsibility	Time Frame	Completed
	<p>3. What do we need to know about the portfolios?</p> <ul style="list-style-type: none"> • Teach group to formulate questions. • Share with whole group. • Standardise common questions in the numbering. • Consider key ideas from May 1 • Each group to write a detailed plan. 			
	4. Each group to finalise their action plan	All Teachers (4 groups)	May 15 - May 29	
	5. Staff Meeting No. 4 Each group will present a typed up Action Plan. Photocopy for each other staff member. Ask other staff for feed back.	All Teachers (4 groups)	May 29	
	6. Each group carries out their Action Plan.	All teachers	No staff meeting on June 5. By June 19	
	7. Analyse data from the surveys.	CHA	Monday June 24 9:30 am.	

ABLE 1996

Our main goal is for all teachers to have knowledge and understanding of the following and to be demonstrating them in practice -

- ~ linking planning to the national and school curriculum objectives
- ~ incorporating assessment into their planning
- ~ integrating assessment into the teaching and learning process
- ~ using a wide range of assessment techniques
- ~ balancing validity and reliability with manageability
- ~ recording assessment outcomes appropriately (i.e. trialing the Profile Book set up in 1995)

A secondary goal is to continue to refine the sample folders with the long term view of having an effective Portfolio system in place. Reviewing its purpose will be the first important issue to address.

To date, the school has -

- ~ completed the baseline survey
- ~ used this information and feedback from staff to establish some goals for the year
- ~ read sections from 'Policy to Practice' and the Education Gazette article.
- ~ held a Teacher-Only-Day (see programme)
- ~ discussed and designed assessment tasks to incorporate into a Social Studies/Health Unit. These units are currently underway. An analysis and review of these assessments is already planned.

: ACTION PLAN

Objective	Action	Responsibility	Time Frame
1. To address concerns raised by staff in the review	1. Feedback to staff at a staff meeting <ul style="list-style-type: none"> - the importance of addressing concerns - the importance of making ongoing changes/ refinements - making recording decisions at the planning stage of a unit of work: what will be recorded and where - recording progress for each child even if the objective hasn't been fully achieved - using planning folders for recording any assessment that will not go directly into the Tracking Files (no Roll Books?) - avoiding duplication - using the Tracking Files for generating a class overview (as well as the other way round) 		Tuesday August 5
	2. Follow up the above in: <ul style="list-style-type: none"> - Senior Staff Meeting - Syndicates 	Senior Staff All staff	Wed 6 August now until end of term
2. To find effective storage boxes for the Tracking Files for each teacher	1. Gather samples 2. Select the best box	All staff	By 16 August

∴ **ACTION PLAN continued...**

Objective	Action	Responsibility	Time Frame
3. To review then finalise the English Scheme and Tracking File sheets	1. Set up a Coordinating team	/Coordinating Team Convenor Coordinating Team Coordinating Team In syndicates - all staff Coordinating Team	by 13 September
	2. Set the dates for two ½ day workshops		
	3. Convenor to contact Kerry to seek her assistance at the 2 workshops		
	4. Write a handwriting form for the Tracking File		By 13 September
	5. Use the 2 workshops to prepare a draft English scheme and review/revise the Tracking File Sheets in English		
	6. Discuss drafts with the staff		
	7. Finalise the scheme and Tracking File sheets at a ½ day workshop		By 11 October
4. To design an assessment framework for Opaheke school that dovetails - the Tracking File - the Portfolios - Reporting to parents that meets needs, is manageable and <u>avoids duplication</u>	1. Copy of Bayswater Questionnaire to Murray Questionnaire to parents	Jan/Assessment Team (M,S,B)	By 9 August
	2. Analysis of questionnaire	Assessment Team	By 13 September
	3. Information on portfolios	Jan to the Assessment Team	August
	4. Design the framework	Assessment Team/Jan	23 August - Workshop 8 October - school visit
5. To review then finalise the Maths Scheme and Track-	As per No. 3	Maths Coordinating Team	Term 4

APPENDIX E

**Sample of fax form used to set up
the workshop agendas**

FACSIMILE

TO ERDC

FAX NO. 443 9381

FROM *SCHOOL* GROUP 3

DATE *31.5.96*

Topics/Activities we would like included in the Development Group Workshop on Friday 14 June are:

Portfolios - what to include

- what not to include

- standardising procedures

- asking - what? who? when

- DO WE NEED PORTFOLIOS AS WELL AS PROFILES?

WHAT'S THE PURPOSE OF THEM?

*- ANALYSING STAFF FEEDBACK - WE NEED SOME TIME
TO DO THIS.*

*- WE STILL HAVE SOME STAFF RESISTANCE TO OUR
PROFILES - HELP!!*

Everyone to bring actual examples and their diaries.

Please fax this to ERDC before **FRIDAY 31 MAY**

Thanks.

APPENDIX F

**End-of-study lead team evaluations
used to review the year's work and
to plan for the following year**

IN-SCHOOL DEVELOPMENT

Thinking about school-wide practices

What changes have been made?

Formalised Planning Sheet includes:

- *Essential Skills*)
- *Achievement Objectives*) *Whole school planning format*
- *Learning outcomes*)
- *Assessment tasks*)

2 year overview cycle for coverage of all strands

Portfolio -

- *goal setting format*
- *annotation sheets/styles*
- *mid year/end of year curriculum summary sheets*
- *guidelines → contents*

Have the changes been successful? Why? (or why not?)

Yes

- *purposes/intentions of portfolios understood and clearly defined by all (as checked by 'wave of change')*
- *links to planning and other documentation identified and carried through*
- *direction and focus for planning (via overview)*

What were the key influences in bringing about those changes?

- * *to eliminate double-handling of assessment data*
- * *to meet the needs of all audiences*
- * *ability to critically reflect upon changes via assessment contract*
- * *staff realisation for need for change in line with curriculum documents*

What have been the greatest difficulties in making the changes? Why?

- * *lack of familiarity with new curriculum documents*
- * *planning to Achievement Objectives/Learning Outcomes annotations*
- * *time constraints*
- * *staff changes*

Thinking about classroom teachers' attitudes and practices

What changes have been made?

- * *Focused annotations, planning*
- * *Greater sense of accountability to each individual child*
- * *View assessment as integral part of Learning/Teaching cycle
- not an extra/add on*
- * *Less double handling*

Have the changes been successful or not?

Definitely

What were the key influences in bringing about those changes?

- * *Purpose of samples in portfolios were not clear*
- * *Annotations were undefined and did not indicate achievement of learning outcomes*
- * *Assessment often only occurred at a removed period of time from teaching/learning cycle*
- * *Teachers were recording same information 3 x*

What have been the greatest difficulties in making the changes? Why?

- * *Time*
- * *Only one staff member has met these changes with vigorous resistance. All other staff have been more than willing.*

IN-SCHOOL DEVELOPMENT

Thinking about school-wide practices

What changes have been made?

Complete re-vamp of the assessment procedures and practices

Development of an individual student profile

*Development of a school prescription/learning outcomes.
Document for all teachers.*

*Reviewed and re-developed the programme for reporting to
parents.*

Have the changes been successful? Why? (or why not?)

*The development of the profile has but we will not know what the
teacher response will be until trialed in 1997 and worked on.*

*Positive feedback on reporting to parents and the fact that a lot of
doubling up of assessment records is going to be reduced if not
eliminated.*

What were the key influences in bringing about those changes?

These changes came about through -

1. Discussions with other schools
2. Discussions among ourselves (as a team)
3. Discussions with Jan Hill
4. Feedback from staff
5. Reading information - books, articles
6. Viewing other schools' profiles and assessment methods
7. Courses on assessment run by Massey and one run by K Smythe

What have been the greatest difficulties in making the changes? Why?

- * *Coming to some agreement as to what method would suit our school*
- * *Coming to some agreement about how to use our portfolios (which we have not yet reached).*
- * *Changes in staff, overseas teachers*

Thinking about classroom teachers' attitudes and practices

What changes have been made?

Teachers seem positive about the profile

Planning is now gearing up to the document by focussing more specifically on achievement objectives. Also teachers are designing more specific assessment tasks.

Have the changes been successful or not?

Unsure until 1997 as far as the profile is concerned

Classroom teachers attitudes and practices are more focussed on individual children and this includes student self-assessment.

What were the key influences in bringing about those changes?

Staff meetings - showing assessment doesn't have to be time consuming.

What have been the greatest difficulties in making the changes? Why?

Pre-conceived ideas and mind-sets

Organisation - making time.

IN-SCHOOL DEVELOPMENT

Thinking about school-wide practices

What changes have been made?

1. *Individual tracking files were introduced for all staff to use from beginning of Term 2.*
2. *Teacher only day was held Day 1, Term 2 to lay the groundwork for implementing change.*
3. *The English scheme has been revised to reflect consistency between the curriculum document, the scheme and the tracking file.*
4. *A move has been made for teachers to move beyond just making context-based evaluative statements to comments that indicate actual achievement.*
5. *Storage boxes for the files were purchased for each class.*
6. *Front page of file detailing standardised test results has been modified.*
7. *Some school-wide units have been implemented as an outcome of the assessment contract with a specific focus on assessment and linking it to unit objectives.*
8. *Changes in planning format introduced that will lead to better assessment practices.*

Have the changes been successful? Why? (or why not?)

All changes have been implemented successfully apart from the context-based evaluative statements which require further and ongoing development.

What were the key influences in bringing about those changes?

Jan, as our facilitator, was the major influence in bringing about a change in understanding and practices of assessment.

The ongoing contact and goal setting was invaluable.

What have been the greatest difficulties in making the changes? Why?

Staff turnover

Attitudes to 'more change'

The school has had too much involvement in off-site contracts (4).

Thinking about classroom teachers' attitudes and practices

What changes have been made?

They are using the curriculum documents to a greater degree to guide them in their planning.

Assessment practices are now an integral part of planning.

Teachers are more aware of the variety of assessment techniques and practices.

The positive effects of planning whole school topics.

Other teachers are now prepared to seek individual guidance in assessment techniques.

Have the changes been successful or not?

Yes.

What were the key influences in bringing about those changes?

The Assessment Committee with the help of Jan.

What have been the greatest difficulties in making the changes? Why?

New staff members

Too many contracts at the same time

Resistance to major change.

IN-SCHOOL DEVELOPMENT

Thinking about school-wide practices

What changes have been made?

- * *Use of curriculum documents*
- * *Consistent tracking system*
- * *Focused planning/assessment tasks*
- * *Focused resource buying*
- * *Budget planning and prioritising*
- * *Uniform planning sheet throughout - ensures reference to essential skills*
- * *Introduction to Benchmark Portfolio in English*
- * *Review of - Report Form*
- *Pupil discussion report*
- *2 year curriculum document*
- * *Form 1 to 4 - development of Junior Assessment*
- Years 7-10 sheet
- * *Education of parents to new system*

Have the changes been successful? Why? (or why not?)

YES

- * *Teachers are aware of ongoing learning*
- * *Meeting staff and children's needs*
- * *Whole school have similar planning and assessment. Less stress and change of planning focus when shifting levels (new teachers)*
- * *Accountability to children*
- * *Better use of funding for resources*
- * *Less teacher stress - not adding more - better time management*
- * *Accountability for staff*
- * *Parents are informed of children's needs/strengths*
- * *Where to next? approach*
- * *System in place for new staff*

What were the key influences in bringing about those changes?

- * *Inconsistent planning/assessment throughout schooling*
- * *Repetition of topics*
- * *Teachers weren't accountable for strand coverage*
- * *Doubling up of resources - resources being used once*
- * *Learning was not monitored individually in an ongoing way*
- * *No link from Junior/Middle/Senior school in skill development*
- * *Teacher overload for limited value - not always meeting children's needs or accountability*
- * *Lack of use of essential skills*
- * *Parents wanted accurate information on where their child was at.*
- * *Wanted user friendly system*
- * *Make secondary teachers more accountable - for assessment; for meeting individual needs*
- * *Allow for co-operative planning*
- * *Time*

What have been the greatest difficulties in making the changes? Why?

- * Resistance to change from staff to new system
- * Re-organising resources - budgeting
- * Parent education of new system
- * Cost of printing - time to put folders together
- * Storage of folders
- * Monitoring assessment - time to do it
- building it into the term's programme

Thinking about classroom teachers' attitudes and practices

What changes have been made?

- * *Cooperative planning*
- * *Excellent use of resources*
- * *Teachers more focused on curriculum documents and coverage of essential skills*
- * *Pre-planned assessment tasks*
- * *Less stress*
- * *Ongoing assessment throughout term*
- * *Keen, enthusiastic because they know what they are doing and where they are going*
- * *Accountability*

Have the changes been successful or not?

- * *Some (2) teachers still not using the schoolwide planning sheet.*
- * *Teachers still need reminding that they need to refer to curriculum documents.*

2PD35

back to your point about coming to an understanding, don't forget that this is not going out on it's own. It goes out inside a portfolio which backs it up and contains the understanding. The context for learning and what has been covered has been listed down the right hand side there. It also comes with an accompanying sheet which you haven't got, and it comes as part of a whole document. It should not be used singly on its own.

levels

2PD35

The huge part of our job is to educate our parents isn't it, in whatever system we have got.

parent ed.

Exactly, community education is vitally important.

2PD35

If there is a problem, take maths, you have a child that is kind of on level three and kind of on level four. What do you do with that child.

This is a continuum. We struggled with this, it is a very good point, well done. It brings me to that point, that perhaps we haven't illustrated it very well in the way the sheet is set out. We did wrestle with this and there will be some modifications. Where you place your indicator shows whether you are between 3 and 4.

levels

2PD35

J- Fill one in for us.

I will fill it in at random.

So this is meant to be an arrow here.

2PD35

If you had all the levels out and just had the line and go down, you could do that, it would look better, it would look like a continuum.

J- Those things (putting a tick or cross) actually make a difference. Kay and I have started collecting a whole lot of work because we kept seeing mastery and not confident etc and we think there is a mix of words, different kinds of comparisons being made etc.

descript:

IK3

This ignorif it - pl/copier talk, he syndicate
 angst betwe - decisions were made as
 a syn - Poly. honesty

IK3

if not honest sets kids up for disaster
 parents misinformed
 better to deal with parent
 quiet sooner or later honesty

P. now a working document.

but Send up best portfolios eg - for open night
 visitors.
 ∴ conflict.

parent to do it
 betwe p. goes
 home.

Encourage kid to be honest
 if didn't - set up to fail.

IK3

Ch huge ownership now.

stud. union

Documentation to support it so important

Before no know. of portfolios.

easy to confuse with sample b/w

IK3

Discussion v. lpt - v. formative to me. sch. prn
 - that happened at mtgs

pl/copier room.

helped to clarify.

listened & discarded. I'm tired.
 for me.

reporting

③

> Reporting on achievement

- est purpose
- using tchr docs
- schools' responsibilities

parent ed.
interesting
many samples
honesty

- traditions
- reviewing the needs of parents
- parent education
- being truthful/honest

Report
tchrs

Computerisation = presentation
= training.

> Role of tchr = types of assessors
= autonomy
= rel. with learner (honesty)

teachers' individual styles and systems

- attitudes
- practices
- duplication

autonomy V consistency
helping V testing culture
↕
rel. of tchrs with students

NOT bodies?