WELLINGTON COLLEGE OF EDUCATION LIBRARY
A876149B

PROGRAMME EVALUATION AND EDUCATIONAL POLICY DEVELOPMENT: AN EXPLORATION

PATRICIA HELEN JACKMAN

2005

T 371.260993 JAC
STACK

Victoria
UNIVERSITY OF WELLINGTON
Te Whare Wānanga
o te Ūpoko o te Ika a Māui

LIBRARY

81735

Programme Evaluation and Educational Policy Development: An Exploration

By

Patricia Helen Jackman

A thesis submitted to
Victoria University of Wellington
in partial fulfillment of the requirements for the degree of Master of Education.

2005

Victoria University of Wellington
Te Whare Wananga o te Upoko o te Ika a Maui

ACKNOWLEDGEMENTS

My grateful thanks go to my supervisor, Dr Geraldine McDonald. Her wisdom, challenges, and support kept me focused and resourced. Her calm approach and gentle insistence balanced my "bull at a gate" attitude.

Thanks to Dr Joanna Higgins who provided support, valuable ideas, avenues to pursue and people to talk to.

To my husband, Dr Tony Jackman, thank you for reading and rereading as my thoughts became more focused. Tony also kept the household ticking over, leaving me free to apply myself to the task.

Appreciation goes to the Ministry of Education personnel who gave of their time and knowledge. All have a tremendous passion for what they do and a commitment to better education for all.

The opportunity for study and the support given by the Wellington College of Education is greatly appreciated. Without this I would not have had the means or confidence to contemplate further study.

Family, friends and colleagues have given me much support and encouragement. This has eased the frustrations and anxiety sometimes experienced during the process.

ABSTRACT

Programme evaluation in education began as a form of public inquiry and has developed into a tool for informing policy development. This process has accompanied the government's focus on outcomes rather than outputs and the current global demand for accountability. In recent years there has been an increase in the letting of contracts by the New Zealand Ministry of Education for the production of evidence to support educational policy and this has included the evaluation of programmes designed to improve teaching and raise student achievement.

The study reports the historical development of programme evaluation and the different schools of thought which have evolved. It outlines the management of formative programme evaluation within the Ministry of Education's Research Division and describes Rist's approach to policy making, used in the Numeracy Development Project. Two large-scale programmes, the Strengthening Education in Mangere/Otara (SEMO) Project and the Numeracy Development Project, are discussed as examples of initiatives involving programme evaluation. The results of both have informed policy and have been extended more widely. The relationship between research and programme evaluation is discussed with reference to the Performance-Based Research Fund.

TABLE OF CONTENTS

Page

Acknowledgements i

Abstract ii

Table of Contents iii

Chapter One

1.0 Introduction 1

1.1 Programme Evaluation 2

1.2 Policy 3

1.3 The Role of the Ministry 4

1.4 Issues Between Programme Evaluation and Policy Development 4

1.5 Recording the Outcomes of Programme Evaluation 5

1.6 My Background 6

1.7 Chapter Organisation 7

Chapter Two

2.0 The Development of Programme Evaluation 8

2.1 The Age of Reform in Educational Evaluation 1800-1900 9

2.2 The Age of Efficiency and Testing 1900-1930 10

2.3 The Tylerian Age 1930-1945 11

2.4 The Age of Innocence 1946-1957 11

2.5 The Age of Expansion 1958-1973 12

2.6 The Age of Professionalism 1973-1983 13

2.7 The Information Age 1984 - the Present 14

2.8 Summary 15

Chapter Three

3.0 Issues in Defining Programme Evaluation 16
3.1 Resources 16
3.2 Research 16
3.3 Evaluation 17
3.4 Qualitative versus Quantitative Method 18
3.5 Programme Evaluation in Education 19
3.6 Standards and Ethics 20
3.7 Roles of the Evaluator 21
3.8 Use of Programme Evaluation 23
3.9 Utilisation-Focused Evaluation 24
Criticisms of Utilisation-Focused Evaluation 25
3.10 The Political Nature of Programme Evaluation 25
3.11 Policy Development 26
3.12 Criticism of the Funding of Educational Research 27
3.13 The Relationship between Research, Evaluation, and Policy 28
3.14 The New Zealand Government 29
3.15 Perspectives on the Relationship Between Policy and Research 29
3.16 Summary 31

Chapter Four

4.0 Methodology 33
4.1 Research Question 33
4.2 Theoretical Framework 33
4.3 Sources of Data 33
4.4 Interview Sample 34
4.5 Questionnaire 34
4.6 Method 35
4.7 Ethics 35
4.8 Limitations 36
4.9 Summary 37

Chapter Five

5.0 Programme Evaluation and Policy within the Ministry of Education 38

5.1 Background 38

The Structure of the Ministry of Education and Policy Determination 38

Ministry of Education Research Record 41

The Structure of the Research Division of the Ministry of Education 42

Information Collected from the Interview Questions 42

5.2 Findings 43

5.3 Policy development strategies favoured by the Ministry of Education 52

5.4 The New Zealand Government's Analytic Frameworks 53

Best-Evidence Synthesis 55

Best-Evidence Use 56

Programme Logic 57

5.5 Summary 59

Chapter Six

6.0 Strengthening Education in Mangere/Otara 60

6.1 Description 60

6.2 The Problem 60

6.3 Ministry of Education Response 62

6.4 The Programme Evaluation 64

The Structure of Schools in Otara 66

The Three-way Partnership Involving Aspects of Literacy 66

Boards of Trustees as Partners 67

Early Childhood Primary Link (ECPL) 67

Picking up the Pace 67

Māori Education in Mangere/Otara 68

Initiative Achievements 69

6.5 Summary 69

Chapter Seven

7.0 Exploring Issues in Mathematics Education 70

7.1 Description 70

7.2 The Problem 70

7.3 Ministry of Education Response 70

7.4 Intervention Phases 72

7.5 Programme Evaluation 73

7.6 Policy Development within the Numeracy Development Project 75

Initiative Achievements 76

7.7 Summary 76

Chapter Eight

8.0 Summary and Discussion 78

8.1 Origins 78

8.2 Outputs and Outcomes 80

8.3 The OECD Report 81

8.4 Policy Targets 81

8.5 Two Ministry of Education Programme Evaluations 82

The SEMO Project 83

The Numeracy Development Project 84

8.6 Outcomes in Policy 85

8.7 Evaluators 87

8.8 Programme Evaluation and the Performance Based Research Fund (PBRF) 88

8.9 Conclusion 89

References 90

Tables

I The Journey of Programme Evaluation 8

II Adapted from the Programme Logic Matrix Structure 58

Figures

1. Research Model 38

2. Ministry of Education Structure 38

3. Lines of Communication within the Ministry of Education 40

4. Phases of the SEMO Intervention 63

5. Phases of the Numeracy Development Project 73

Appendices 97

One: Consent Form 98

Two: Letter of Introduction 99

Three: Schedule of Interview Questions 100

Four: Number Framework 102

CHAPTER ONE

1.0 Introduction

The *Dominion Post* 24th April 2003, in an editorial headed "Are we getting value for money?" commented that "... a fundamental tenet of spending in the modern public service is that it be measured not on 'outputs' but 'outcomes'....". Outputs are the programmes that departments, ministries and other agencies are asked by Ministers to provide; whereas outcomes are the results that the policies actually deliver. It is the outcomes, therefore, that need to be accounted for and the process by which their success is measured involves evaluation.

In this era of accountability for all management actions, O'Neill (2002) says in her Reith Lectures, that educationalists, as part of the public sector, have often felt that they are not trusted. This lack of trust is generally expressed as a failure of central authorities to delegate full control over the establishment of standards. New Zealand schools are said to be self-governing but in practice they are required to respond to requirements and targets of accountability for standards over which they do not have full control. For example, they must comply with detailed health and safety requirements and increasingly complex employment legislation in the development of which they have not been involved. Accountability has therefore become associated with measurement and evaluation by outside agencies.

The call for accountability has also resulted in an increase in the demand for systematic empirical evaluation of the effectiveness of government programmes. This has, in turn, brought new challenges involving how to use the findings from the process of evaluation, and how to avoid possibilities for their misuse. Education is not a single unified process and evaluation is not simple. The evaluation of educational programmes, particularly programmes designed to improve some aspect of schooling, is a process that takes place in what are often very diverse situations in schools. Judgment of the use of such evaluations poses the problem of the applicability of the findings to schools of many different kinds.

1.1 Programme Evaluation

Programme evaluation is an American term to describe the process of judging whether a programme is valued and effective. A programme can be anything from a new system of management to a full-scale research project. Some New Zealand examples are:

A project that delivers services to clients as in:

Genetic Change (Charteris, 2002) in which evaluation is used to quantify the genetic merit of animals and rank them.

A mass media campaign aimed at specific target population as in:

Evaluation of the national "Push Play" campaign in New Zealand: creating population awareness of physical activity (Bauman, McLean, Hurdle, Walker, Boyd, van Aalst & Carr, 2003).

A new law as in:

Evaluation of Compulsory Breath Testing in New Zealand (Bailey, 1995).

Evaluation and research are closely related in that evaluation generally involves the processes of research and research frequently involves some form of evaluation.

But, whereas in programme evaluation the focus of the study is already defined by others, either in the form of an existing programme, or outlined in a commissioned study as in the many contracts let by the New Zealand Ministry of Education, the focus of research may be determined by a researcher pursuing a personal interest.

Like research, evaluation has its different schools of thought and different purposes. Patton (1997), a leader in the field of evaluation, sees programme evaluation as a formative process; as "the systematic collection of information about the activities, characteristics, and outcomes of programmes to make judgments about the programme, improve the effectiveness of it, and inform decisions about future programming" (p. 23). His special contribution to the field has been the idea of *utilisation-focused evaluation* to categorise evaluation that is done for, and with, specific intended users and for specific intended uses. Stake's (1998) definition of programme evaluation is "a technological subdivision of educational research which includes objective and subjective inquiry in the disciplined search for programme quality" (p. 203). Therefore programme evaluation goes beyond the simple establishment of facts in the search for "best practice" or "what works" to use some current phrases.

1.2 Policy

Morris and Fitz-Gibbon (1978) state that originally programme evaluation was seen as a contributor to long-term programme effectiveness and improved decision-making and was synonymous with measuring *goal attainment*. Today programme evaluation has evolved to play a role in the *shaping of policy*.

According to Stone (1997), policy may be defined as a "rational attempt to attain objectives" (p. 37) while Rist (1998) sees policy-making as multidimensional and multifaceted and thus programme evaluation is only one of the components that can contribute its formulation. Smith (2002) argues that the complexity of policy development has left both the policy makers and citizens less able to understand the issues and to see how their actions may affect present conditions. To develop effective policy, Irving (1992) contends "there is a need for policy advice that demonstrates a thorough and up-to-date knowledge and understanding of research literature, drawing on relevant research evidence and information both locally and internationally, and evaluating what can or cannot reasonably be deduced from it" (p. 1). Programme evaluation, therefore, can be thought of as a way to clarify the issues affecting proposed changes in educational methods and institutions and to evaluate the quality of existing research. As Irving (1992) noted, policy analysts need to be familiar with research, research methodologies and data analysis. However, to make use of research for policy purposes, programmes introduced to make improvements in education need to be evaluated and, thus there should be a close working relationship between policy analysts and evaluators/researchers. This is especially so in the development and funding of long-and short-term research, in planning, implementing and evaluating interventions, and in analysing findings. Irving sees this relationship between analysts and evaluators/researchers as imperative if the government is to receive policy advice that addresses immediate and local issues and which draws on an understanding of relevant national and international research and results from the evaluation of specific programmes.

1.3 The Role of the Ministry

The New Zealand Ministry of Education is responsible for the development of policy in accordance with the agenda of the government of the day and it has within its structure, both research and policy sections. However, their relationship between the sections is more than a structural one and relies on the willingness of members in each group to reconcile

professional differences in designing programmes to meet the government's desired outcomes and the educational needs of the student population.

In 1998 the New Zealand Government adopted the goal that "By 2005, every child turning nine will be able to read, write and do maths for success" (cited in Higgins, Parsons & Hyland, 2003). McMahon (2002) reports that the demographic predictions for the year 2040 show that New Zealand will, by then, have a white minority population. Moreover, he emphasises that the present Māori and Pasifika students, who make up the bulk of those who are failing in the system, will be the work force in 2040. This calls for both educational programmes to suit these students and for methods of evaluating their effectiveness. Following its evaluation, the intervention, funded by the Ministry of Education, called Strengthening Education in Mangere and Otara (SEMO) may provide a model contributing to raising student achievement nationally.

The recent emphasis on evidence-based policy and the development of best evidence syntheses[1] indicates the way in which the government is trying to address the issues raised by Irving in 1992. To be seen as well spent the money spent on the evaluation of a particular programme should be capable of informing decision making.

1.4 Relationship Issues between Programme Evaluation and Policy Development

The field of programme evaluation is still in the process of development and Smith (1980) identifies a range of issues relating to it. He argues that more effort is needed in improving the relevance and utility of evaluations and their cost effectiveness[2] and suggests that expanding the role of consumers and stakeholders[3] in evaluative activities could help achieve this. This is close to Patton's arguments. Smith (1980) also notes the need to increase the role of programme evaluations in providing equal educational opportunity, an issue relevant to the government's goals. The evaluator's ability to deal with moral and ethical problems is, he believes, also in need of strengthening and he has also urged the

---

1 Best evidence synthesis – "The purpose of the synthesis is to contribute to ongoing, evidence-based and evolving dialogue about pedagogy amongst policy-makers, educators and researchers that can inform development and optimize outcomes for students in New Zealand schooling " (Alton-Lee, 2003, p.v).

2 Cost effectiveness – the extent to which one programme, project, or instructional material produces equal or better results than competitors that cost about the same amount of time, effort and resources; or the extent to which an object produces the same results as competitors but is less costly.

3 Stakeholders – individuals or groups who may affect or be affected by programme evaluation.

creation of a professional identity of programme evaluation to guide the future of evaluation and the development of an empirical basis for evaluation practice. Programme evaluation should be based on more than simple judgment.

Programme evaluators may encounter multiple values and conflicting interests at evaluation sites. The field of programme evaluation may also involve disputes over what is research and what is evaluation and a tension between qualitative and quantitative methodology. Gorard (2002) believes "the over-used dichotomy between qualitative and quantitative is a reason why sometimes evaluation evidence does not, and cannot, support reported conclusions and yet is passed by peer-review" (p. 141).

Programme evaluation can link the "outputs" and "outcomes", mentioned in the editorial cited at the beginning of this chapter, when it is used to judge the worth of educational programmes aimed at raising student achievement. The educational programmes themselves, as in the case of SEMO, are supported by government funds, often on a very large scale, and as Robson (1993) points out, justification of expenditure calls for some kind of objective evidence that the product is valuable. Therefore, programme evaluation is important on several grounds. It has the capacity to answer whether a programme "works", whether money is being spent wisely, and whether a better way to achieve educational outcomes can be found.

1.5 Recording the Outcomes of Programme Evaluation

In reviewing evaluation as a professional activity Boulding (1980) observed that one of the factors which distinguishes a mature and secure profession from one that is immature and insecure is that the former systematically records and analyses its history. The point is particularly relevant to New Zealand where systematic recording of the development of programme evaluation has yet to be undertaken. Such a project would be too ambitious to attempt in this thesis but in the absence of systematic recording this thesis explores a number of topics relating to research/evaluation and policy.

The research question that provided a central focus for the study is:

> *What part does programme evaluation play in the relationship between research and policy within the New Zealand Ministry of Education?*

However, when I worked my way into my study I found that the answer depended upon a number of other issues, such as the origins of programme evaluation, its relationship to research, its use in determining policy, and its outcomes in changes in education. As a consequence the work developed into small studies of the separate topics all of which must be considered exploratory.

The thesis, therefore, sets out to explore:

1. the origins of programme evaluation in New Zealand.
2. its beginnings in existing research.
3. the role of the Ministry of Education in deciding upon and commissioning evaluations.
4. the linking of programme evaluation to policy development within the Ministry of Education.
5. practical examples of the work of the Ministry in setting up and evaluating programmes can be found in Strengthening Education in Mangere and Otara (SEMO) project, and in the Numeracy Development. These will be described and their policy outcomes discussed.

1.6 My Background

For the past 35 years I have been involved with primary school education both in New Zealand and, for four years, in the United States of America. Initially my position was as a classroom teacher and my career path took me to senior management positions in low-decile schools. I joined the Wellington College of Education staff as an adviser in 1995. My initial position was as a Mathematics Contract Facilitator and in 1996 I began the role of Rural Adviser concentrating on school management and administration in the small schools in the Wellington region. During the last four years I have moved into the area of formative assessment, delivering the Ministry of Education Assess to Learn (AtoL) Initiative.

In my pursuit of tertiary qualifications I have been studying programme evaluation and the processes by which this is carried out in New Zealand

In life, and particularly in education, whenever there is someone evaluating something, the evaluator is seldom neutral. As an Education Adviser I am often involved in putting into effect Ministry of Education policies. It has not always been clear how such policies have been arrived at and, if they involve programme evaluations, exactly how they have been evaluated. Reflecting on this, I became interested in educational programme evaluation use in New Zealand but there is not a well-established field of enquiry on which to base my efforts. My rationale for dealing with the issues surrounding programme evaluation, its development, its relationship to policy determination and its results, was that, although there have been studies of the separate components there did not appear to have been a syntheses of how they were linked in New Zealand. The two programme evaluations described were recommended by the Ministry of Education as examples of both, effective practice in informing policy development and how the initiatives within them have been extended nationally.

1.7 Chapter Organisation

The structure of the exploration is as follows. Chapter Two outlines a brief history of evaluation from a global and New Zealand perspective. Chapter Three explores issues in defining programme evaluation, and Chapter Four discusses the methodology used to develop the exploration. Chapter Five briefly outlines the structure and work of the Ministry of Education related to policy and evaluation, the current work in programme evaluation and the New Zealand government's preferred strategies for evaluative studies of programmes. Chapters Six and Seven describe the SEMO and Numeracy Development Projects. Chapter Eight provides a discussion of the information gathered on programme evaluation, its uses, and its effect on research and policy within New Zealand.

At the end of each chapter a summary is provided and the important points are further developed in the discussion in Chapter Eight.

CHAPTER TWO

2.0 The Development of Programme Evaluation

What people understand evaluation to be, and how they go about it has proceeded through what Madaus, Stufflebeam and Scriven (1983) referred to as "ages", or what Guba and Lincoln (1989) called "generations". Each "age" or "generation" builds on the past but does not necessarily displace it. The progression of "ages" and "generations" is summarised in Table I.

The term "programme evaluation" in education serves to distinguish the evaluation of students in classrooms from the evaluation of intervention programmes aimed at improving learning and teaching, and policy initiatives.

Table I

The Journey of Programme Evaluation

Created from information in:

Madaus, Stufflebeam and Scriven (1983)

The Evolution of Evaluation		
Age of Reform 1800-1900	Evaluations took the form of Official Commissions of Inquiry	Evaluation as a form of public inquiry
Age of Efficiency & Testing 1900-1930	The idea of scientific/quantitative management in education becomes predominant	
Tylerian Age 1936-1945	Both quantitative evaluation and its technical aspects were further developed in this period	
Age of Innocence 1946-1957		
Age of Expansion 1958-1972	The beginning of profound changes that would see evaluation expand into an industry and a profession	
Age of Professionalisation 1973-1983	The field of evaluation emerged as a profession quite distinct from research and testing.	Evaluation as a field of its own emerged with prescribed rules and methods both quantitative and qualitative and integrated.
Age of Information 1983 - present		Programme Evaluation has gone from merely generating findings to generating knowledge

Guba and Lincoln (1989)

The Coming of Age of Evaluation		
First generation: Measurement 1895-1930	Measure success with various tests	Evaluator is a technician – creates or uses existing instruments to measure any possible variable
Second generation: Description: 1930-1967	Explain discrepancies between the objectives and results	Evaluator is a describer – describes patterns of strengths and weaknesses
Third generation: Judgment 1967-1979	Judge the programme's relative performance	Evaluator is a judge – makes a judgment on the programme's merit and worth
Fourth generation: Negotiation 1979-present	Collectively reconstruct the programme in both quantitative and qualitative ways, integrating both.	Evaluator is a mediator, acts as a negotiator, facilitator and catalyst for social change

2.1 The Age of Reform in Educational Evaluation 1800-1900

The late 19th century was a time when systems of education were established. New Zealand, for example established a national system in 1877 (Department of Education, 1978). As problems arose, Government-appointed commissions carried out what today might be called programme evaluations. One of the topics, which led to inquiry, was rates of failure and a commission in France, of which Alfred Binet was a member, studied "retarded children". In 1908 Binet (Wolfe, 1973) developed a test to identify children who exhibited learning problems and this led to the testing of intelligence.

Madaus et al., (1983) describe the early situation in the United States. In Boston schools in 1845, with what is thought to be the earliest formal attempt to evaluate performance in America, Samuel Gridley Howe began the long tradition of student test scores being used as a measure of school effectiveness, and the comparison of schools. Together with Horace Mann, the father of American public schooling, Howe attempted to establish differential effects from the student data. The data were used to eliminate headmasters who opposed their view on the abolition of corporal punishment. According to Madaus et al., (1983), this was the first clear politicisation of evaluation data.

Between 1887 and 1898, Joseph Rice (Madaus et al., 1983), conducted what is generally acknowledged as the first formal education programme evaluation in America. The evaluation was of spelling and looked at whether a lot of time or a short time was spent learning spelling lists. Despite the fact that the findings showed little difference it resulted in changes in spelling teaching strategies.

This saw the beginnings of the experimental design approach, followed in the late 1800s by the accreditation movement. The former involved the first use of comparison between experimental and control groups (Madaus et al., 1983).

In New Zealand the first school established for Māori was at Rangihoua in 1816. There were not enough pakeha children or sufficient interest in education to warrant a pakeha school (Department of Education, 1978).

2.2 The Age of Efficiency and Testing 1900-1930

The terms *appraisal of learning*, *examining* or *testing* were used prior to 1930. In America the idea of scientific management became a powerful force in the theory of educational administration. The use of surveys followed the making of a connection between instructional outcomes. Evaluations were sometimes seen as "muck raking" employed by a few local people who invited outside experts to expose "defects" and propose "remedies" (Madaus et al., 1983). Test data continued to be the base of evaluations and these were used to compare educational systems

In America between 1920 and 1930 Evaluation Institutes were established. These could be considered as the precursors of American university centres dedicated to evaluation. The evaluations tended to deal with localised questions but were characterised by standardisation, systemisation, and efficiency. In 1929 Ralph Tyler (Madaus et al., 1983) was appointed director of the Bureau of Educational Research of Ohio State University and began to advocate the significance of objectives, within evaluative analysis.

In New Zealand a number of reports and commissions were completed: for example, in 1908, the Hogben Report, *How New Zealand Schools Differ from those Overseas,* and in 1930, the Atmore Report, *Educational Reorganisation in New Zealand* (Department of Education, 1978). These, however, unlike American attempts to evaluate education were not driven by formal testing.

2.3 The Tylerian Age 1930-1945 (WW11: 1939-1945)

Different writers present different interpretations of the practice of evaluation. Patton (1997) contended that the closest thing to evaluation in the 30s was a few jobless academics writing programme histories and yet Ralph Tyler is often referred to as the Father of Educational Evaluation. Tyler (Madaus et al., 1983) coined the phrase "educational evaluation" which meant assessing the extent to which valued objectives had been achieved as part of the instructional programme. Tyler then conceptualised evaluation as a comparison of intended objectives with actual outcomes. Tyler, and his colleague Smith, took this concept a step further with the comparison of intended outcomes against actual outcomes (Madaus et al., 1983).

Through an initiative of America supported by the Carnegie Foundation in 1933 the New Zealand Council of Educational Research (NZCER) was established to promote the cause of research and investigation of education in New Zealand. C. E. Beeby was the first Officer and later Director of the NZCER (Department of Education, 1978). He became Director of Education. In 1939, the famous Fraser-Beeby statement appeared in the report of the Department of Education to Parliament, "Every person, whatever his level of academic ability, whether he be rich or poor, whether he live in town or country, has a right, as a citizen, to a free education of the kind for which he is best fitted, and to the fullest extent of his powers". The initiatives being implemented in New Zealand schools in the post war era had equality of opportunity as their priority (Department of Education, 1978) and this statement became an objective in Tyler's sense.

2.4 The Age of Innocence 1946-1957

This could be called The Age of Ignorance. Poverty and, in the United States of America, despair went unnoticed by those unaffected. There was an expansion of education but with things in abundance no one was held to account. Data collected was used to justify expansion rather than to judge and improve programme quality (Madaus et al., 1983). The need for standards for the control of evaluation became apparent. Government agencies had not yet become involved in the evaluation of programmes.

In New Zealand the use of committees as a primary source of reviews and reports continued. A first Research Officer, Dr E G Jacoby, was appointed to the Department of

Education in 1948. He gained an international reputation for his work in the field of school population projections (Department of Education, 1978).

2.5 The Age of Expansion 1958-1973

This "age" was the beginning of changes that saw evaluation expand as an industry and into a profession dependent on taxpayer monies for support. However, despite President Kennedy and Lyndon Johnson pouring money into programmes with the aim of equity, good funding, and new technology, evaluators were not very successful in finding acceptance. Also in America, the 1964 Elementary and Secondary Education Act was amended to include specific evaluation requirements in which evaluators were required to shift from the realm of theory to that of practice and implementation (Madaus et al., 1983). Evaluators became aware of the complexity of the social context of programmes and of the increasing number of social issues the programme implementers faced in achieving quality outcomes. In the 1960s however, social research was not accepted as a valid measure of programme quality and so did not influence the development of policy. In Weiss's (1972) view, even evaluations carried out by the United States government showed no clear or direct effect on decision-making.

Growing disquiet over the failure of standardised testing as an ideal tool for programme evaluation saw Phi Delta Kappa set up the National Study Committee on Evaluation in 1971. Stufflebeam, Eisner, and Scriven, (Madaus et al., 1983) and others developed new approaches to programme evaluations. A number of important evaluations resulted in negative findings raising questions about the value of programme evaluation and this set the scene for the next "age." It was felt that formalised programme evaluation was a fad and would soon be proved ineffective and so disappear (Madaus et al., 1983).

America has supplied the foundation of programme evaluation for the world where it grew in strength after 1965 when mandated for social programmes. However, Patton (1997) says that under the surface and behind the scenes, a crisis of utilisation, both in non-use and misuse of evaluation findings was building up.

In 1970, The New Zealand Department of Education's Research and Planning Unit was established and the Unit took part in the International Evaluation of Educational

Attainment (IEA) research for the first time (Department of Education, 1978). This started New Zealand's participation in a series of studies the results of which were an evaluation of standards in a number of different curriculum areas. The judgment of New Zealand's standing was assessed in relation to the scores of other countries.

2.6 The Age of Professionalism 1973-1983

Evaluation began to emerge in America as a distinct profession related to, but quite distinct from, research and testing. The field was fraught with confusion, anxiety and animosity and evaluators questioned whether they should be researchers, testers, administrators, teachers or philosophers (Madaus et al., 1983).

In an effort to professionalise their field which had little status and no political clout, a number of evaluation journals was started, and two societies were established: The Evaluation Research Society (academically-oriented) and the Evaluation Network (practitioner-oriented). Universities offered evaluation courses and in some American States evaluators could apply for certification (Madaus et al., 1983).

Evaluators began questioning the quality of evaluations with the consequence that meta-evaluations were carried out and evaluation standards developed in 1982 (Patton, 1997).

There was substantial professional development in evaluation but polarisation between quantitative and qualitative approaches reflected ideological differences. A widely accepted definition for programme evaluation at this time was one designed and conducted to assist a potential audience to judge and improve the worth of an educational object, but many studies did not conform to this definition and, in some cases, evaluators opposed it.

In 1974, Patton developed Utilisation-Focused Evaluation, which focuses on intended use by intended users, a focus that is paramount at all stages of programme evaluation from the beginning to the end (Patton, 1997).

In 1973 the New Zealand Department of Education's first Research and Planning Unit Bulletin was published and the largest study of education in New Zealand history took place in the Educational Development Conference, which involved over 50,000 people.

Evaluation studies of pilot schemes in nursing in 1973 saw the first use of the term *evaluation* in the Department of Education publication titles (Department of Education, 1978).

In 1974 the Department of Education Research and Planning Unit was renamed Research Statistics Unit; in 1978 this became The Research and Statistics Division. In 1978, the then Minister of Education, the Hon L Gandar, convened a two-day conference on educational research. This was the first to be held in 40 years (Department of Education, 1978). The New Zealand Association for Research in Education (NZARE) was founded in 1979.

2.7 The Information Age 1984 – The Present

In 1984 the American Evaluation Society and the Evaluation Network merged to form the American Evaluation Association. According to Patton (2001), evaluation associations had formed in many countries and national, and international evaluation conferences ensured, and continues to ensure, rigorous debate and sharing of ideas as the world looks to programme evaluation to bring order from chaos (Patton, 1997). In 1994 the Educational Programme Evaluation Standards were adopted.

By the 1980s in New Zealand it was clear that the government had not met the equity goals on which the education system was supposed to be based. In the late 1980s and continuing into the 1990s, New Zealand underwent a radical restructuring of its economic and social institutions. The broad thrust of these reforms was not unique to New Zealand and they reflected a worldwide trend to accept the discipline of market forces. The New Zealand Treasury was involved in every facet of governmental decision-making. In 1987, Treasury's brief to the incoming government was significant in the development of educational policy in New Zealand (The Treasury, 1987). According to McDonald (1993), "Treasury had made extensive use of research, in particular, sociological research to make a case for the application of free market principles to all levels of education. The document signaled the start of a campaign to alter the nature of the government's responsibilities for the provision and management of education" (p.1). Middleton, Codd and Jones (1990) believe the Treasury advice was a strategic response to a fiscal crisis, which was impacting on the government's ability to fulfill its role in social service delivery.

Since the 1980s there has been an increasing tendency in New Zealand to move to forms of research/evaluation where the focus of the investigation was determined by the funding agency, rather than the predilection of the researcher/evaluator or the priorities of their institution (Livingstone, 1989). Livingstone (1989) documented a number of trends in the relationship between research and policy within the Department of Education. While there was more effective communication between research and policy, a new constraint in the relationship appeared to be financial.

2.8 Summary

Programme evaluation, and more specifically, educational programme evaluation, has evolved from the late 19th century to today. Methodological shifts occurred from a primary emphasis on quantitative data (often student test scores), to qualitative methods involving stakeholders in the evaluation process. Programme evaluation has evolved from a form of public inquiry into a field of its own with prescribed rules and methods. More recently, it has been argued that the role of the evaluator should change from that of an expert to a negotiator who facilitates the process of programme evaluation for different stakeholders.

In New Zealand, educational programme evaluation was a late development. It was not until 1973 that the Ministry (Department) of Education's first Research and Planning Unit was established.

The current state of educational programme evaluation in New Zealand is outlined in Chapter Five.

CHAPTER THREE

3.0 Issues in Defining Programme Evaluation

It is not always easy to distinguish between programme evaluation and research, and there are disputes over their boundaries. Therefore, although this review deals mainly with issues around programme evaluation there are brief discussions of research and of policy development.

3.1 Resources

Two texts frequently referred to in connection with programme evaluation are *Utilisation-Focused Evaluation: The New Century Text* (Patton, 1997), and *Evaluation Models: Viewpoints on Educational and Human Services Evaluation* (Madaus, Scriven, & Stufflebeam, 1983). Both texts reflect changing viewpoints on evaluation design and its purpose.

The literature on programme evaluation is scattered across several disciplines. Some writers, including Patton (1997) and Madaus et al. (1983), have summarised the major trends and concepts. The World Wide Web also offers information on programme evaluation. The field has its own journals such as: *Evaluation; Studies in Educational Evaluation; Educational Evaluation and Policy Analysis*; and *Evaluation and Programme Planning*. The ProQuest – Education Complete database and the Educational Resources Information Centre (ERIC) contain studies relevant to programme evaluation.

3.2 Research

Stoney (2000) defines research as "the collection, analysis and interpretation of information undertaken in a systematic manner" (p.1). It is primarily concerned with increasing the knowledge base and enhancing understanding. According to Patton (1990), its purpose is "to generate theory and discover truth" (p. 12). In the opinion of Hume and Bryce (2001), "a researcher has an important function as a public intellectual with responsibility to contribute to debates on matters of public policy … and take account of the agendas and priorities of government but will not be determined by them" (p. 330).

The OECD (2001) categorises research into three conventions:

> basic, strategic and applied. Basic research is not tied to any specific practical goals, but is undertaken primarily to acquire new knowledge of underlying phenomena. Strategic research operates between basic and applied, with a longer time horizon and broader goals than the latter. Applied research is original investigation directed primarily towards a specific practical aim or objective. Evaluation studies are a prime example of this. (p. 14)

It is the impression of the OECD (2001) that "the bulk of educational research in New Zealand is concentrated at the applied end… the balance between different types of research is a salient issue for R&D policy" (p. 14). Within applied educational research, action research is increasingly important as teachers are encouraged to take a greater role in the evaluation of new programmes and to undertake study as part of their professional development. Action research "explicitly and purposefully becomes part of the change process by engaging people in the programme or organization in studying their own problems in order to solve those problems" (Whyte, 1989, cited in Patton, 1990, p. 157).

3.3 Evaluation

Evaluation is not restricted to either education or programmes. However, evaluation is specific in focus whether the focus is a programme, policy, or activity. *The Shorter Oxford Dictionary* (Little, Fowler & Coulson, 1973) says that to evaluate is to "work out the value of, to find a numerical expression for, or to express in terms of the known". The 1994, United States of America Standards for Educational Evaluation (McNamara, 1998) define evaluation as "the investigation of merit or worth". Evaluation is both backward and forward-looking. It is analytical, using both quantitative and/or qualitative data to explore activities and issues. Cronbach and Suppes (1969) (cited in Patton, 1997) believe the difference between research and evaluation to be the difference between conclusion-oriented and decision-oriented inquiry.

Although scientific research is generally described as being undertaken to discover new knowledge, test theories, establish truth, and generalise across time and space, the

distinction between research and evaluation is by no means hard and fast. It is quite possible for programme evaluation to reveal new knowledge. Research can involve the provision of an evaluation, and moreover, evaluation findings can be research based. Zungia (1990) (cited in Shragge, Church, Fontan, & Lachance, 1999) contends evaluation is, above all, an integral part of all planned intervention – a reflection on action.

However, just as evaluation is becoming recognised as a field related to scientific research but with its own standards and methods, Shragge et al. (1999), say that, more and more, the notion of science is being questioned regarding its objectivity, its methods, and its political purpose.

Clearly, there are various views on evaluation reflected in a continuum of "traditional" science, towards various alternative methods of inquiry. However, the definition of evaluation, adopted in this thesis is the systematic application of social research procedures for assessing the design, implementation, and utility of interventions such as programmes, projects, or instructional material. As discussed later this is a definition that fits the use of evaluation in informing decision-making and the development of policy.

3.4 Qualitative versus Quantitative Method

There are two broad methods of collecting data in any research or evaluation. These range from solely quantitative scientific methods, to solely qualitative methods. Programme evaluation has increasingly used the strengths of both quantitative and qualitative methods. It is interesting to note that the New Zealand Ministry (former Department) of Education projects of 20 years ago (e.g. the Learning in Science Projects) used a mix of quantitative and qualitative methods in their evaluation phases.

Patton (2002) states, " one of the greatest breakthroughs in the evaluation field is the end to the qualitative-quantitative debate" (p.2). This appears to be the opinion of many evaluators as they recognize that mixed methods have much to offer.

Denzin and Lincoln (2000) define qualitative research as the "studied use and collection of a variety of empirical materials that describe routine and problematic moments and meanings in individuals' lives" (p. 3). According to Denscombe (1998) qualitative research

tends to be associated with words as the unit of analysis while quantitative research tends to be associated with numbers as the unit of analysis. Qualitative methods enable the evaluator to study selected issues in depth and detail while quantitative data interpretation involves determining correlation or significant difference, both of which are based on mathematical probabilities for which absolute proof is not a possibility.

Patton (1990) contends that, "because qualitative and quantitative methods involve differing strengths and weaknesses, they constitute alternative, but not mutually exclusive strategies of research. Both qualitative and quantitative data can be collected in the same study" (p. 14).

3.5 Programme Evaluation in Education

In any discussion of educational programme evaluation it is important to distinguish between the evaluation of students and the evaluation of programmes affecting students. The evaluation of students focuses on finding the level of achievement of a particular student or cohort in relation to a school, national, or international standard. The evaluation of programmes affecting students is focused on raising student achievement by ensuring the quality of the programmes being implemented, including their effectiveness for their stated purpose. In general the purpose will be raising student achievement.

Patton (1997) states programme evaluation has its roots in the social sciences as a pragmatic process probing the actions of governments on behalf of the public. Its purpose is to assess the effects or effectiveness of something, typically some innovation or intervention: policy, practice or service. It is trans-disciplinary as well as being a discipline in its own right and has developed a problem-solving mode that is client centred, often confidential, and targeted at carefully defined issues and questions. Patton (1997) writes, "Programme evaluation is undertaken to inform decisions, clarify opinions, identify improvements, and provide information about programmes and policies within the contextual boundaries of time, place, values, and politics" (p. 24).

Rossi (1989) (cited in Shragge et al., 1999) sees evaluation as the *systematic application of social research procedures in assessing* the conceptualization, design, implementation, and utility of social programmes. Robson (1993) states that evaluation is a study that has a

distinctive purpose and is not a new or different research strategy but is essentially indistinguishable from research in terms of design, data collection techniques and methods of analysis. This is not a universally accepted opinion. In Pawson's (2001b) view, evaluation differs fundamentally from basic research, primarily in the *purposes* of data collection and the standards developed for judging the quality of outcomes. However, it is Robson's (1993) opinion that there is no widely agreed definition of programme evaluation.

Patton (1997) believes debate around approaches to programme evaluation is necessary but that there should not be a return to only one approach. He reminds us that a parochial practice is one that repeats the same thing over and over again while a pluralistic and cosmopolitan practice is one that adapts evaluation practices to new situations. The one non-negotiable aspect of programme evaluation is that it is based on systematic data collection.

Who should benefit from programme evaluation? Should it reflect and amplify the voices of those who are subject to it or should it serve the interests of those who commission the evaluation? Shragge et al. (1999) argue that programme evaluation emerges from a combination of situations: people; politics; history; context; resources; constraints; values; needs; interests and chance.

3.6 Standards and Ethics

To maintain the reputation of any profession it is imperative that those practicing within it follow the standards and ethics set down. In the United States of America the Evaluation Research Society created Programme Evaluation Standards in 1982, but on the Society's merger with the Evaluation Networks in 1984, none of these standards and guidelines were officially adopted. In 1992 a Joint Committee on Standards for Educational Evaluation (McNamara, 1998) was established to develop evaluation standards as the base guidelines of ethics, standards and principles to be adhered to. These were adopted in 1994 under four criteria: Utility; Feasibility; Propriety and Accuracy. The five principles of: Systematic Inquiry; Competence; Integrity/Honesty; Respect for People; and Responsibilities for General and Public Welfare (McNamara, 1998) support the standards which Shragge, et al. (1999) argue have been fundamental to the development of evaluation within the profession of education but can be applied in all areas of programme evaluation.

3.7 Roles of the Evaluator

It is clear that definitions of programme evaluation are influenced not only by the context and purpose of the evaluation, but are also dependent on the beliefs and values of the evaluator. According to Robson (1993) there is no dissension regarding the need to critically assess the functioning of educational programmes, the issues are more with who does this, in what way, and for what purpose.

Evaluators are normally contracted to carry out an evaluation by an agency or institution. There is an issue as to what extent it is professionally possible and economically wise for evaluators to express independent beliefs and values when their livelihood may depend on future contracts. There are also problems in that evaluators may have to respond to restrictive conditions often aimed at meeting unrealistic deadlines resulting in a situation that can lead to a failure to meet professional, ethical, and quality standards.

Stufflebeam (1994) and Scriven (1991) believe the evaluator renders a judgment about merit and worth, while Patton's utilisation-focused evaluator functions as negotiator between all intended users. However, in Chelimsky's (1995) view, the evaluator's chief purpose is to tell the truth to people who do not want to hear it.

In Coe's (2002) opinion the prior expectations of the evaluators can strongly bias the outcome of the evaluation but most evaluators are not necessarily aware of this tendency in themselves, or of the power it has to distort results. According to Scriven (1991), evaluators normally submit reports with generally positive bias; they have a tendency to turn in more favourable results than is justified. Patton (1997) acknowledges also, that evaluation is an intrinsically sensitive activity where there may be risk of revealing inadequacy. He expresses concern that the evaluator's intention may be misconstrued and findings misused, or even ignored. Because we all have selective perception, Patton (1997) argues that the very action of a person taking in information distorts reality.

There is no doubt that programme evaluation is complex and an evaluator needs to be able to appreciate the viewpoints of others and be able to find ways to design an evaluation that incorporates diverse interests, including their own. However Scriven (1991) cautions evaluators against over reliance on personal judgment in resolving value conflicts.

Shragge et al. (1999) feel the approaches chosen by evaluators are most importantly expressed in whose questions are addressed and, therefore, what criteria are used to make judgments about programme quality. Evaluators have a variety of approaches to a programme evaluation.

Evaluator	Focus on the needs of:
Campbell (1972) and Greene (1994)	policy makers
Scriven (1993)	programme consumers
Patton (1997)	on-site programme administrators and board members
Stake (1995)	on-site programme directors and staff

Moreover, some evaluators are committed to approaches based on whom the evaluation will be useful to:

Evaluators	Useful to:
Cronbach (1982)	consumers
House (1993)	the under privileged
Greene (1994)	the programme's target groups
Fetterman (1997)	the disenfranchised

It is Patton's (2002) belief that "evaluators need to be not only methodologically competent, they also need to be skilled at situational analysis, bring cultural sensitivities to their work, be politically sophisticated about the ways in which data and methods are used in the knowledge age, and understand how evaluation intersects with politics at all levels" (p. 3). These attributes are additional to those required by a researcher but this is seldom acknowledged in the research community.

A vexed question is whether the evaluator needs to be an expert in the field being evaluated. Scriven (1991) and Robson (1993) agree that you do not have to know about the particular field to be able to evaluate a programme. However, the evaluator, while a professional in his own field, should develop some knowledge of the context and content early in the evaluation process.

Beal (2003) believes educational evaluation can sometimes threaten those who are implementing the programme being evaluated, for example, the teachers, as they believe it is they who are being evaluated. It is important that there are guidelines to limit negative repercussions.

The rationale for all approaches presumes there is an audience interested in the information and its interpretation. Often this audience is primarily the institution that commissioned the programme evaluation. In reality evaluation studies are frequently done for the accountability purposes discussed in Chapter One so that the funding institution/agency will be recognised as effective and that the findings will substantiate and support anticipated action, or action already taken, and thus justify or validate further funding.

3.8 Use of Programme Evaluation

The use of programme evaluations is an issue identified by Patton (1997) and many others. Stake (1967) appears to have been the first person to consider the connection between methods and use. His "responsive evaluation" focuses on identifying and addressing the concerns of all the stakeholders in a programme, and examines those concerns in terms of programme improvement.

King and Pechman (1984) contend that use of programme evaluation involves the intentional and serious consideration of information by an individual with the potential to act on it. The concept of usefulness, however, depends upon the perspectives and values of the observer. Two major influences on the extent of the use of programme evaluation are political considerations and the personal factor. The personal factor emerges when some people take direct and personal responsibility for getting findings to the right people; only then, says Patton (1997), can evaluations have an impact.

The 1994 Standards for Educational Evaluation (McNamara, 1998) make it clear that it is the evaluator's responsibility and challenge to ensure use of the results. Patton (1997) and Stufflebeam (1994) see utilisation as vital while Stake (1995) believes utilitisation is not fundamental to the concept of evaluation. Scriven (1991) contends that use depends on the conjunction of availability of results, public concern, and bureaucratic receptiveness - being involved in the process can often be as useful as the formal findings that emerge.

Chelimsky (1997) suggests that the purpose of the evaluation conditions the use that can be expected of the evaluation findings. This use of the findings generally involves expenditure. Patton (2001) believes that how evaluations are used affects the spending of billions of dollars to address societal problems in the United States of America but Brandl (1994) cautions that more money spent on evaluation does not mean higher quality or better evaluative results.

3.9 Utilisation-Focused Evaluation

Patton (1997) talked of a crisis existing in programme evaluation because of the limited use of the findings. His response was to develop utilisation-focused evaluation in which the evaluator facilitates judgment and decision making by intended users and stakeholders rather than acting as a distant, independent judge.

Utilisation-focused evaluation is a process for helping the primary intended users and stakeholders select the most appropriate content, model, methods, theory, and uses for their particular situation. Patton defines two types of use; firstly the use of formal findings to make changes, and secondly the process knowledge gleaned informally by the participants. The latter involves individual changes in thinking and behaviour, often involving organisational and programme changes in procedures and changes in culture that occur among those involved as a result of learning. The right way to do an evaluation is the way that will be most meaningful and useful to the evaluators and intended users involved.

Utilisation-focused evaluation, by its very definition, focuses on the end use from the beginning of a study, and attempts to enhance its eventual impact long before the final report is produced. It offers a philosophy of evaluation and a practical framework for designing and conducting evaluations. According to its proponent, Patton (1997), examples from the inception of utilisation-focused evaluations circa 1978 prove that the evaluation will be used if the foundation for use is properly prepared. This is a problem-solving/formative approach calling for creative adaptations to the changed and changing conditions. It is not a technical approach that attempts to change the conditions to meet the evaluator's needs. Chapter Seven discusses a Ministry of Education programme evaluation, the Numeracy Development Project, which has a formative design.

Whether the expected type and degree of use has been achieved is frequently the focus of the evaluator's evaluation of their own work. This leads to the field of meta-evaluation, which is an evaluation of the evaluation. Patton (1997) says "a meta-evaluation is necessary so that stakeholders have credible, independent review of an evaluation's strengths and weaknesses" (p.143). Meta-evaluation is a common form of monitoring the outcome of evaluation methods in other disciplinary fields, such as holistic environmental assessments.

Criticisms of Utilisation-Focused Evaluation

Patton has his critics. Stufflebeam (1994) and Scriven (1991) raise objections to Patton's (1974) coining of the term Utilisation-focused Evaluation. They believe it is damaging to include in a definition of evaluation any additional concepts. Evaluation is evaluation. Helping people to help themselves is not a fundamental goal of evaluation according to Stufflebeam (1994) and Scriven (1991). Extraneous goals cause confusion, just as there is role confusion when one moves beyond being an evaluator to take on other roles such as a trainer. They believe that the integration of the data collection with the development of the programme contaminates the data and threatens its validity and a loss of independence by the evaluator can undermine neutrality. In a worst-case scenario, such loss of neutrality can spiral into dishonesty, corruption, and data distortion. The argument seems to revolve around evaluator objectivity.

3.10 The Political Nature of Programme Evaluation

In discussing Patton's (1997) Utilisation-focused Evaluation, Waters (1998) points out that stakeholders do not have equal power to influence the programme evaluation and, nor do they have equal protection from the consequences of that evaluation. She believes there can be no doubt that programme evaluation is influenced partly by political forces, and in turn, has political effects.

Cronbach (1982) and Patton (1997) also believe programme evaluations are necessarily intended to serve a political function. In their view society should innovate but that the evaluation of the success, or otherwise, of the innovation, is necessary. Whose interests are served and how interests are represented in a programme evaluation are critical concerns in a society with increasingly disparate value systems. In House's (1993) opinion, programme

evaluation is a combination of both political and scientific authority applied to practical decisions and actions.

Shragge et al. (1999) define programme evaluation as "political" because it can be used to exercise power. The question of who has power is determined by the context in which the evaluation takes place. Greene (1994) points out that any programme evaluation is integrally entwined with political decision-making about societal priorities, resource allocation, and power. According to Campbell and Ng (1988) (cited in Shragge et al., 1999), programme evaluation is one of the techniques used in the administration of state funds to ensure that the interests of the dominant group are successfully (re)asserted over those of a subordinate group.

Floden and Weiner (1978), in discussing the relationship between evaluation and government processes, emphasise the impact of evaluation on discrete decisions made by public managers. Discussion of educational programme evaluation frequently acknowledges the very political environment within which the discussion is taking place. In the words of House (1993) programme evaluation inevitably becomes a political tool in decision development as it affects "who gets what."

3.11 Policy Development

According to the British Government Cabinet Office (cited in Nutely, Walter & Davies, 2002b), policy development is "ultimately about delivering outcomes – desired change in the real world" (p. 8). It is argued that educational policies during the 19th and 20th centuries were strongly influenced by theories (such as Herbert Spencer's Social Darwinism) that were based on the belief in survival of the fittest. Such assumptions informed policies of differentiation in educational provisions for different races, genders and abilities. Post-war educational policies had a dual focus: the rights of individuals to achieve their potential, and perceived social needs.

Tomorrow's Schools (Ministry of Education, 1988) set out a new education policy for the New Zealand Government. It stated that the purpose of the Ministry of Education was to provide policy advice to the Minister and to oversee the implementation of national policies.

Today's New Zealand education policy is driven by the goals of the present government:

> *Reducing systematic underachievement in education and reducing disparity.*
>
> *Building an education system that equips New Zealanders with 21st century skills.* (Ministry of Education, 2003a).

In pursuit of their goals The British Government Cabinet Paper (cited in Nutely et al., 2002b) expects policy-makers to better use research and evidence to focus on policies that will deliver long-term goals. They believe better links between research and policy depend on the two communities finding points of exchange at more than a product level.

3.12 Criticism of the Funding of Research in Education

There is literature on the subject of the funding of research in education. The OECD (2001) report for example, expressed some concern about the Ministry of Education being the main funder of contracted educational research. They see a problem in that there is not a wider set of research funders, no autonomous research council and very few foundations ready to provide resources for educational research. This situation could result in limited ability to build capacity within the research community and development of expertise and ability. It also limits the ability to look beyond applied research topics.

The government has acknowledged the need to build capacity within the research community by the introduction of Performance-Based Research Funding (PBRF). According to the University of Otago (2003), this fund is:

> a new mechanism for research funding in the tertiary sector. Funding for research is no longer allocated to institutions according to student enrolments. Instead, research funds are allocated through the PBRF according to the quality of the research produced in each institution. The system will take into account the quality of researchers, research degree completions and external research income (p. 1).

Research is judged to be representative of one of four quality categories: A, B, C, and R (previously D). Quality A gains the highest funding and represents pure academic research. At the applied research end of the continuum programme evaluation attracts the lowest

funding. Through the PBRF tertiary institutions are ranked against each other. The implications of this new system will be discussed later.

3.13 The Relationship Between Research, Evaluation, and Policy

Lomas (2000) contends that policy making is a diffuse, haphazard, and somewhat volatile process. Policy-makers are concerned with products from, and not processes within, the research community. They want assistance with emerging problems upon which policy is currently being formulated. Lomas (2000) describes this research as "a retail store – as if researchers are busy filling the shelves of a shop-front with a comprehensive set of all possibly relevant studies that a policy-maker might some day drop in to purchase. This view recognises neither the breadth of possible studies that could be done, nor the numerous stages involved in choosing which of those studies to do and how to do them" (p.142). It also contrasts with the "tailor-made" product resulting from contracting and commissioning as in the New Zealand Ministry of Education Research Division.

Rist (1998) has said: "So long as researchers presume that research findings must be brought to bear upon a single event, a discrete act of decision making, they will be missing those circumstances and processes where, in fact, research can be useful" (p. 1003). Hume and Bryce (2001) support the view that research into policy should not be reduced to research for policy which would see policy-makers control research.

Lomas (2000) believes more attention is required on both sides in establishing and maintaining ongoing links and more comprehensive communication.

For the most part, the funding of programme evaluation is driven by the needs of policy-makers. According to Pawson (2001a), effective policy process uses evidence from pilot studies and ongoing programme evaluation. However, evaluating possible policy alternatives in advance of formulation of policy is critical but as Patton (1997) says, the issue is how?

3.14 The New Zealand Government

Perhaps, as a possible solution to the relationship between research division and policy sections, the New Zealand Government (2002b), promotes "intervention logic." It is stated that, in the "development of an integrated management system driven by robust outcome information, and change management… systems should be underpinned by good outcome measures, robust intervention logic and the careful alignment of input, output and outcome information" (p. 6). Intervention logic is a tool for results-based management or the careful scrutiny of policy arguments and is used as a basis to justify new and continuing policies. The New Zealand Government (2002a) describes intervention logic as "an evidence-based, systematic and reasoned description of the causal links between outcomes and the department's outputs, and the associated assumptions and risks" (p. 5). It can "take one or more of three basic forms: narrative, flow diagram, or a framework, and can be enhanced by further systematic testing of assumptions implicit in the intervention logic" (p. 2). The New Zealand Government's (2002a) "preferred framework for Intervention Logic is aligned to the version of Programme Logic developed by Funnell (1997)" (p. 6). Funnell's version can be represented in the form of a matrix (see Table II). The strength of this model is that it differentiates between factors either within or outside the influence of the demographics of the target group, competing programmes, or the past experiences of the programme clients.

In the past there has been a lack of interaction between research and policy, and the non-inclusion of stakeholders in the process but more recently the relationship has been formalized by building research and policy functions into institutional frameworks.

3.15 Perspectives on the Relationship between Policy and Research

According to Pawson (2001b) policy making is a conceptual, conjectural and self-revising process in which empirical inquiry (research) cannot make its voice heard amidst the clatter of other, political imperatives on policy making.

As Austin (1988) explained, the Ministry of Education, as a policy and advice structure cannot advise without an adequate evidence base – that is, one cannot devise policy in a vacuum. At face value it would appear reasonable to assume that all types of research would have a role to play in contributing to policy development. Gordon (1993) contended

that administrative reform had worked to exclude teachers from involvement in central policy processes and felt that the reform implementation process was a process of detailed consultation rather than a process of policy development. However, Baker (2000), in reference to the reform, contends that, "these changes have not been accompanied by a systematic plan of research to evaluate the impact of the changes, which could inform ongoing developments" (p. 4).

Irving (1992), writing over 10 years ago, believed that the relationship between research and educational policy was not satisfactory. He saw a need to balance the current policy requirements with the longer-term research requirements not immediately or directly related to current policy development. He advocated that the policy sections of the Ministry should be closely involved in the allocation of funding for research and that good policy advice should demonstrate a thorough and up-to-date knowledge and understanding of research literature and draw on relevant evidence both from within New Zealand and internationally.

It appears that many share the view that policy needs to be informed by research, however, Hume and Bryce (2001) warn that "at the interface between a policy section and a research division, one group has the power to bring about change and the other is more or less impotent; politicians (at least those in office) have power in that they can make sure their intentions come to pass" (p. 347). At best, research can have only a moderate influence on policy formulation and implementation. Ball (1990) and McPherson and Raab (1988) suggest that ideology, bureaucratic systems and the networking of powerful individuals and groups have often more potent influence in shaping policy outcomes than evidence from programme evaluation.

The views expressed by Lomas (2000), Rist (1998), Hume and Bryce (2001), the British Government (1999, cited in Nutely et al., 2002b), Ball (1990), McPherson and Raab (1998), and Pawson (2001a,b) are significant. Each is concerned with the relationship between research, policy, and practice in education and the need for them to be seen as processes and not products. All are developing different approaches to the provision of evidence to inform decision making. As they explore various approaches to this issue the crux of the problem is, whose evidence is counted, and what is counted as evidence.

The discussion in Chapters Five, Six, and Seven will describe the progress the New Zealand Ministry of Education has made in developing processes to strengthen the relationship between research and policy, and ensuring that such processes include all stakeholders.

3.16 Summary

The literature makes it clear that programme evaluation has its roots in the social sciences and has evolved into a profession with its own standards and ethics. Methodologically it was originally dominated by the natural science paradigm, which values quantitative measures, but today the alternative approach in which qualitative methods are valued is popular. Neither of these paradigms is more effective than the other and can be successfully combined but there is still tension between them in the view of some commentators.

The question of evaluator role represents a real issue in programme evaluation. Different types of, and purposes for, programme evaluation call for varying roles for the evaluator which span from rendering judgment about merit and worth of the programme, to the role of a negotiator between all the programme stakeholders.

In response to the crisis of the limited use of the findings of programme evaluation utilisation-focused evaluation has been developed. By its very definition it focuses all decisions from the beginning of a study to the eventual impact of the study. Not everyone agrees on the effectiveness of the utilisation-focused evaluation approach but it is a credible approach to the use issue.

The concept of usefulness of the programme evaluation depends upon the perspectives and values of the observer but two major factors that inhibit or enhance the utilisation of a programme evaluation are political considerations and the personal factor. However, politics is often the overriding factor.

In the interface between research and policy, programme evaluation inevitably becomes a political tool in that it affects "who gets what." However, for educational programme evaluation to inform decision making and policy there needs to be an understanding by

both researchers and policy makers that each needs the other. As a possible tool to develop the relationship the New Zealand Government has introduced the use of intervention logic as a basis to justify new and continuing policies.

CHAPTER FOUR

4.0 Methodology

4.1 Research Question

What part does programme evaluation play in the relationship between research and policy within New Zealand?

4.2 Theoretical Framework

The primary purpose of the study was to carry out research focused on programme evaluation. The research question provides a unit of analysis, which is the relationship between research and policy.

To understand the relationship I needed to know:

What programme evaluation is

How it differs from research

How policy is decided

How evaluation fits into this

How programme evaluation works out in practice

How such programmes inform policy

4.3 Sources of Data

The exploration is based on discussion of various kinds of information.

The analysis of programme evaluation has been discussed in Chapters One and Three, as has its difference from research. It was noted that although there was no final agreement on definition or purpose there was a common core.

How policy is decided was based on the work of the Ministry of Education and interview with some of its representatives.

How programme evaluation worked out in practice was explored in two different Ministry of Education contracted programme evaluations, the SEMO project and the Numeracy Development Project.

4.4 Interview Sample

The interview sample was readily identified within the population of interest. Three key staff from the Ministry of Education Research and Evaluation Team, and the Learning Policy Frameworks Team were invited to take part in semi-formal interviews because they were directly involved in the focus of the exploration, namely programme evaluation, its uses, and effect.

I had met one of my informants in my tertiary study and in my position as Senior Adviser, Wellington College of Education. When approached this informant, who was from the Research and Evaluation Team felt it would be beneficial to have a colleague present at the interview to allow two perspectives on programme evaluation within the Ministry of Education. This proved to be very worthwhile as it ensured broader coverage of the topics discussed.

When carrying out the interviews the questions I had prepared remained largely unasked, because the informants, aware of the focus of the interviews provided the answers without prompting. This fact reduced any bias that could have resulted from my unconscious wording of the questions.

4.5 Questionnaire

In developing the interview topics I reflected on the information presented and asked:

What did I want to know?

What prompts and probes might I need to use to get the information I wanted?

What answers did I expect?

The questions developed were open-ended enough to allow for free-flowing discussion.

The schedule of questions is reproduced in Appendix Three. For the reason given above they should be considered as topics covered rather than as actual questions asked.

4.6 Method

In considering the method by which to carry out research in the development of this exploration a qualitative approach was used for the interviews. Other information came from official documents, general literature on the topic, and information posted on the web.

Qualitative methods have their own approach to collection and analysis of data, the data being produced through interpretation and use. Patton (1990) contends that analysis is difficult because responses are neither systematic nor standardized. Thus, this exploration is a study of discovery - based on inductive analysis rather than preexisting expectations Patton (1990).

4.7 Ethics

I am very close to the situation under study. My position as an Adviser and a researcher is a sensitive one. In my position I advise in the Ministry of Education School Support Services contract, and at the same time, because the Ministry funds much of the programme evaluation in education carried out in New Zealand, I am exploring the Ministry's system.

Ethical approval was gained through the Wellington College of Education's ethics committee. Informed consent for the interviews was obtained from all the informants. The formal consent form and information sheet presented to informants prior to the interview are shown in Appendix Two.

Interviews took place at the Ministry of Education in February and March of 2003. The interview appointment times were negotiated to cause the least inconvenience to the informants. The interviews were of an hour in duration, which was the amount of time set by the informants, and gave sufficient time for the information to be collected. A tape recorder was used to record the interviews, but this failed to work in the second interview and I had to take notes.

The transcripts were returned for checking to those I interviewed. After member checking the necessary changes were made. The informants reviewed a draft of the study areas that relate directly to themselves prior to completion of the thesis. All transcripts will be held until one year after the thesis has been deposited in the library at which point they will be destroyed.

Identities have not been revealed and material from the interview is not accessible to anyone other than myself.

To support the information provided by Ministry of Education personnel that related to programme evaluation, I analysed the Research Division research record of contracts awarded. This is available on the Ministry website.

4.8 Limitations

I initially felt the information I collected would be limited because I carried out only two interviews but this reservation was dispelled during analysis of information from the histories, interviews, discussions, literature review and the two programme evaluations. At this time I found I had too much information. This required my sorting and prioritizing information specific to, or informing, the focus of educational programme evaluation.

Because the tape recorder broke down during the second interview it meant I had to question the adequacy of the notes I took. However, the informant accepted the transcript as a true record of the interview. Rather than trying to write down everything said during the interview I had noted what I viewed as major points. I also began writing up the written notes at the conclusion of the interview while my mind retained the content.

Summary of Procedures

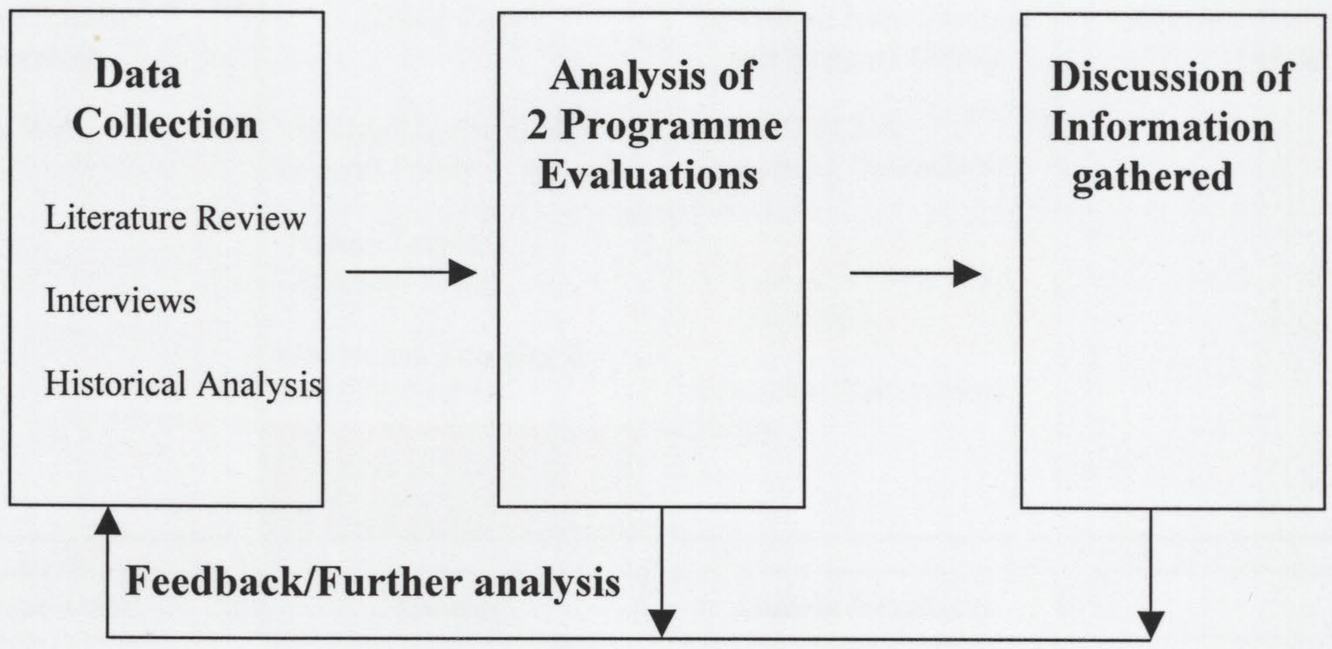

Figure 1. Research Model

4.9 Summary

The research question led to three fields of inquiry; programme evaluation, its relationship to research, and the relationship of both in the area of policy development within the New Zealand Ministry of Education.

The data were collected by the study of documents and by interviews. The empirical information was subjected to qualitative analysis. Ethical procedures were described and limitations discussed.

CHAPTER FIVE

5.0 Programme Evaluations and Policy within the Ministry of Education

5.1 Background

The Structure of the Ministry of Education and Policy Determination

The Ministry of Education has undergone many restructurings post 1989 which increased my interest in its organisation. The following diagram shows where the groups, divisions, sections and teams, fit within the structure of the Ministry as of March 2004.

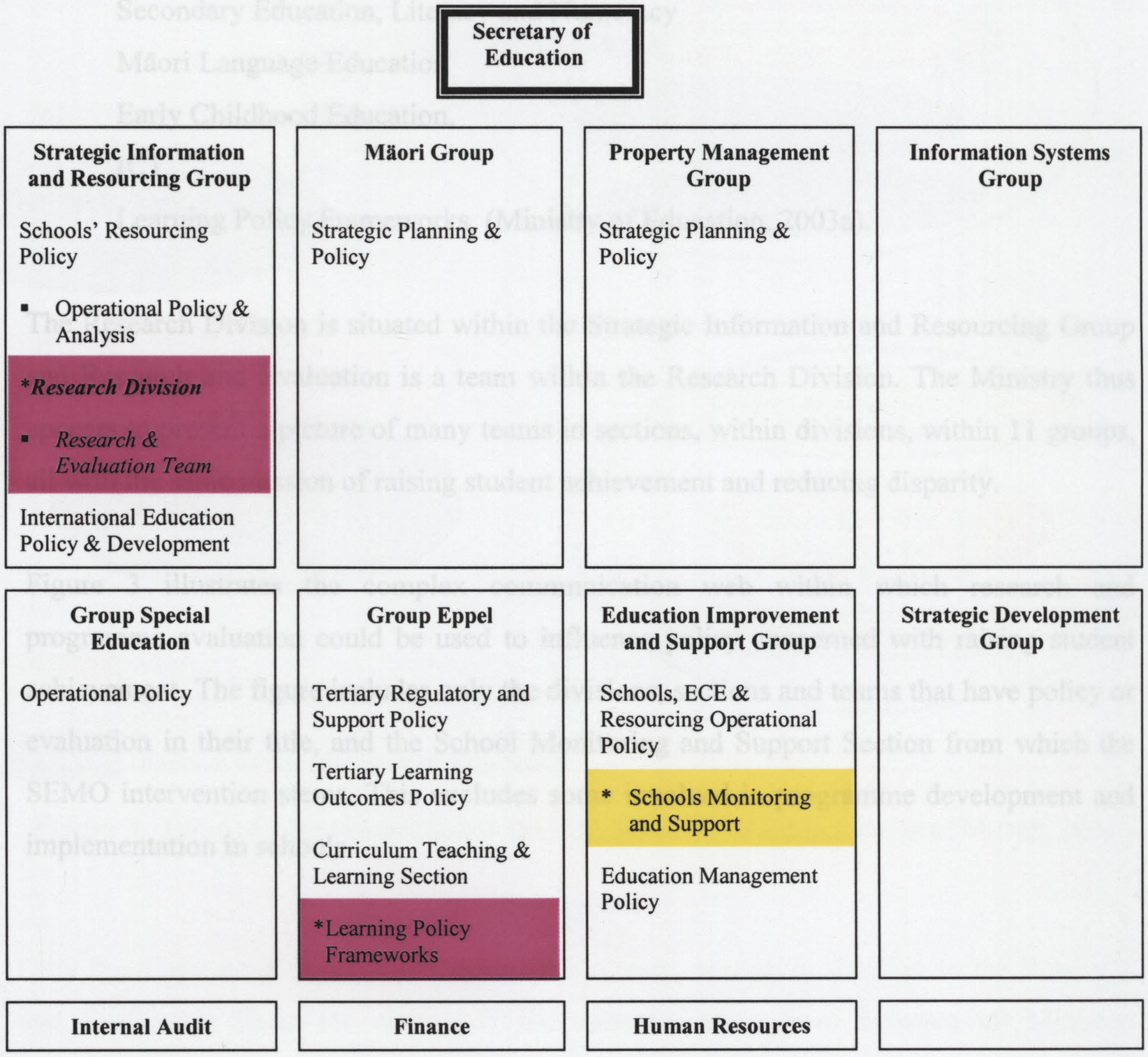

Figure 2. Ministry of Education Structure

Adapted from the Ministry of Education *Statement of Intent* 2003-2008 (Ministry of Education, 2003a).

*Shaded areas are those on which the thesis focuses.

To understand the functions of programme evaluation within the Ministry of Education's Research Division and policy sections. I selected the Learning Policy Frameworks Team as the most appropriate to seek information from. As shown above, the Learning Policy Frameworks Team lies within the Curriculum Teaching and Learning Section of Group Eppel. Within this section there are nine teams all of which focus on programme development and delivery within schools:

Teaching and Leadership Capability
Outcomes and Assessment
Pedagogy and Learning Materials
Secondary Education, Literacy and Numeracy
Māori Language Education
Early Childhood Education,
ICT
Learning Policy Frameworks (Ministry of Education, 2003a).

The Research Division is situated within the Strategic Information and Resourcing Group and Research and Evaluation is a team within the Research Division. The Ministry thus appears to present a picture of many teams in sections, within divisions, within 11 groups, all with the same mission of raising student achievement and reducing disparity.

Figure 3 illustrates the complex communication web within which research and programme evaluation could be used to influence policy concerned with raising student achievement. The figure includes only the divisions, sections and teams that have policy or evaluation in their title, and the School Monitoring and Support Section from which the SEMO intervention stems. This excludes some involved in programme development and implementation in schools.

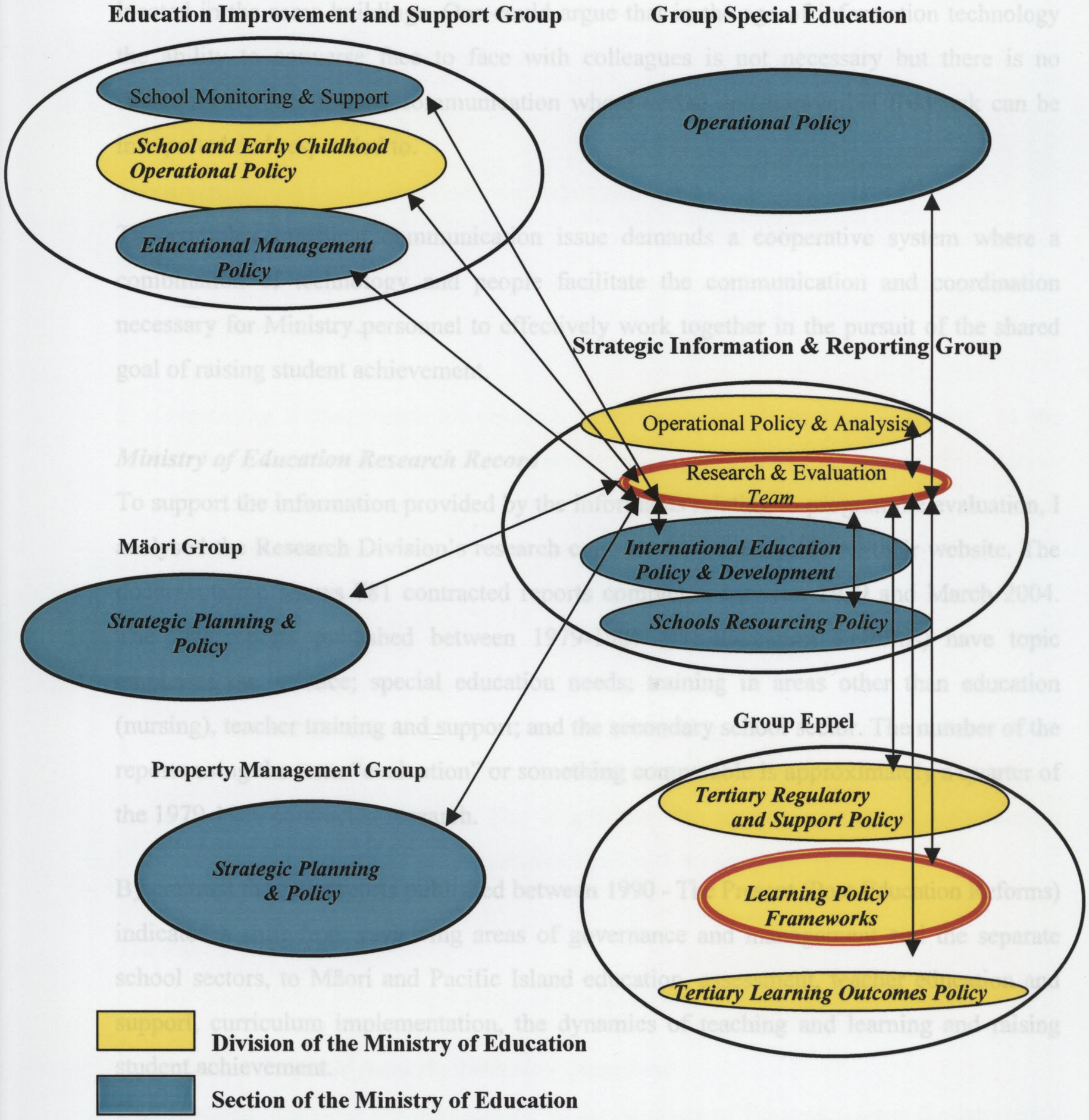

Figure 3. Lines of Communication within the Ministry of Education as at March, 2004

Only the direct lines of communication from a team, section or division to the Research and Evaluation Team are shown. If the communication lines between all Ministry personnel were included in the figure the web created would be unreadable, yet this would represent reality. Therefore questions about intra-communication arise. Adding to the difficulty of communication is the fact that not all groups, sections, divisions or teams are

located in the same buildings. One could argue that in the age of information technology the ability to converse face to face with colleagues is not necessary but there is no substitute for face-to-face communication where verbal and non-verbal feedback can be interpreted and responded to.

The complex logistical communication issue demands a cooperative system where a combination of technology and people facilitate the communication and coordination necessary for Ministry personnel to effectively work together in the pursuit of the shared goal of raising student achievement.

Ministry of Education Research Record

To support the information provided by the informants relating to programme evaluation, I analysed the Research Division's research contract record available on their website. The documentation shows 781 contracted reports completed between 1979 and March 2004. The 280 reports published between 1979-1989 (Pre-Education Reforms) have topic emphases on: science; special education needs; training in areas other than education (nursing), teacher training and support; and the secondary school sector. The number of the reports using the term "evaluation" or something comparable is approximately a quarter of the 1979-1989 contracted research.

By contrast the 501 reports published between 1990 - The Present (Post-Education Reforms) indicated a shift from reviewing areas of governance and management and the separate school sectors, to Māori and Pacific Island education, assessment, teacher education and support, curriculum implementation, the dynamics of teaching and learning and raising student achievement.

The number of these reports using the term evaluation, or something comparable, is almost a half of the contracted research.

The universities undertake over half of New Zealand's educational research with independent consultants contributing approximately one fifth. When I asked my Ministry of Education informants if they felt that the introduction of Performance-Based Research Funding (PBRF) would have implications for the tertiary sector they indicated they couldn't speak for that sector but saw no problems ahead in getting their contracted research completed.

The Structure of the Research Division of the Ministry of Education

Today, the Ministry of Education Research Division is described on the Ministry website as comprising:

The Research and Evaluation Team which has two roles:

1. Working with others in the Ministry to provide them with current, research-based information for use when developing, modifying, and implementing specific educational policies and practice.

2. Overseeing a programme of research undertaken by researchers "external" to the Ministry of Education. The projects researched/evaluated are initiated by the Ministry of Education and funded either by the Research Division's strategic research budget and/or other divisions within the Ministry from where the project stems and/or specific budget appropriation.

The Comparative Education Research Unit (CERU):

Responsible primarily for facilitating and overseeing New Zealand's participation in large-scale international research projects such as The Third International Mathematics and Science Study (TIMSS), The Programme for International Student Assessment (PISA), and the Progress in International Reading Literacy Study (PIRLS).

The Publishing Unit:

Responsible for publishing a range of research reports annually.

Information Collected from the Interview Questions

Unless stated otherwise the information on the Ministry of Education since the late 1980s was gleaned from the interviews with three Ministry of Education personnel. As reported in the methods chapter, although the questions were not asked, the informants provided the answers as they gave their accounts. The questions form the structure of following account although the informants' answers may come from different sections of the intended schedule of topic.

5.2 The Findings

What has been the development of programme evaluation in the Ministry of Education post 1989?

On this topic the informants stated that since the late 1980s programme evaluation has been driven by the increased emphasis on the importance of evidence in the Ministry of Education as "they work to become more skilled about how, when, and where to intervene" (Ministry of Education, 2003a, p. 3). The Government needs information about the effectiveness of current educational spending with a particular focus of this spending on raising student achievement. The Research Division is increasingly accountable to Government and central agencies such as Treasury, Te Puni Kokiri and the Audit Office, in terms of providing evidence through evaluation as to what policies and programmes are working.

What are the significant changes in the approach to programme evaluation post 1989?

As the Research Division meet the demands of increasing accountability they carry out strategic policy research, that helps identify key factors in education outcomes, which feed into strategic policy at the national level. The Division essentially asks three inter-related evaluative questions:

> Are policies and programmes making a difference?
>
> How do we know?
>
> How can we make a bigger difference?

As indicated in the *Dominion Post* April 24, 2003, the New Zealand Government requires Ministries to focus on outcomes. "The ultimate objective of outcome-based management systems is to define, in explicit terms, the results agencies intend to achieve, and to create a management culture that is fact based, results oriented, open, and accountable" (New Zealand Government, 2002b, p. 2). The focus is especially on outcomes affecting diverse students and improving professional capacity to improve achievement for all and reduce disparity.

What is the Research Division's approach to managing ongoing formative programme evaluation?

The Research Division is more outcome focused, and so is doing more formative evaluation. According to Lau (1997) the defining purpose for formative evaluation design is to show stakeholders what is working and what is not, with a particular interest in improving the programme. Lau (1997) states,

> The formative and evolving evaluation design, besides helping teachers, also yields important data about school change and programme effects. The findings include not just the results of certain changed practices, but also information concerning the process itself. Moreover, involving teachers as internal evaluators yields insights that an outsider might miss. These findings about the process are useful in informing the project in making mid-course changes and may be helpful to other projects attempting to make changes in schools through similar efforts. (p. 14)

With regard to a formative approach the Ministry of Education asks many evaluative questions:

> Is there a better way?
> How can we build on the work already done and influence the work to follow to achieve greater coherence and co-ordination?
> What are these bigger questions telling us?
> What key things/principles are consistently evident across a number of evaluations?
> What is the place of ongoing monitoring?

What processes are in place to strengthen the line between research and policy?

In 1996 a new management structure in the Ministry of Education appointed eight new division managers who encouraged a more direct connection between policy and research.

There was a move away from the field-initiated studies, which tended to be small evaluations of policies and initiatives viewed as "happiness quotients" in that little took

place as result of them, and they did not always link with the policy work programme. The focus switched to work that made a more medium term contribution to the Ministry.

In 1999 a Strategic Research Initiative began what was a draft statement of educational research priorities. This was done internally, although nine literature reviews were commissioned from outside sources. These reviews contribute to the draft statement of the Research Division.

The topics of these reviews with the institution or researchers stated in brackets follow:

The impact of family and community resources on student outcomes: an assessment of the international literature with implications for New Zealand. (Stanford University)

Early Childhood Education. (Otago University)

The Effects of Curriculum and Assessment on Pedagogical Approaches and on Educational Outcomes. (University of Waikato)

Influence of peer effects on learning outcomes: a review of literature. (University of Auckland)

Literature review of the effects of school resourcing on educational outcomes. (BERL/Infometrics)

The effects of school governance, ownership organisation and management on educational outcomes. (Rentoul & Posanowski, with Dempster, Fisher, Hosking, Hunter, Pugh and Walford)

Human resources issues in education. (University of Toronto)

Monograph on quality in post-compulsory education. (Education Directions)

Enterprise based education and training: a literature review. (Monash University/ Australian Council for Education Research). (OECD, 2001)

As the reviews of research were completed the summaries and reports were made available on the Ministry of Education website and contributed to the evidence base for the development, on an ongoing basis, of future policy.

In late 2000 the large tertiary institutions in New Zealand came to talk through research needs at a Ministry of Education-convened Strategic Research Initiative Forum. Following the Forum the Ministry discussed what its outcomes meant in terms of research priorities. Consequently a Strategic Information Overview Group was established which focused the research programme on teaching and the dynamics of learning, bilingual and immersion education and tertiary education, and made a deliberate move away from governance issues.

How is programme evaluation funded?

Before 1996, a pool of approximately $1,000,000 per annum in research funding was available through the Research Division of the Ministry. Bids for what the government called "field-initiated" work, that is work not generated by the Research Division of the Ministry of Education, were made to the Research Division at any time. They were assessed and a decision on whether to fund any bid was made. This process had a drawback in that it lacked any formal relationship to policy development. The relationship between the research/evaluation that was funded and the information policy-makers needed was often ad hoc, serendipitous, and not always matching current needs. There was the further issue of not knowing what kinds of bids were about to come in and hence, difficulty for the Ministry in planning ahead.

In the informants' opinion, the three-year electoral cycle of government often leads to evaluators having limited time to carry out programme evaluations, or programme evaluations that may be presented to a government different from the one which requested it, and which will therefore possibly have no impact. Historically, the Ministry had no way of knowing if it was going to have sufficient quality proposals to allow its funding to be spent by the end of the year. Some evaluators realised that if they made their bid in May they were highly likely to get a contract because, at that time, the Ministry of Education usually had plenty of unspent money. The Ministry of Education now has a process for prioritising its research programme and puts out Requests for Proposals for specific, policy-relevant projects.

The Research and Evaluation Team manages on behalf of the Curriculum Teaching and Learning Team some Assessment Funding Pool projects (the pool funds a number of assessment tool projects: The National Education Monitoring Project (NEMP), the Assessment Tools for Learning (asTTle) Contract, The Assessment Resource Banks (ARBs) Contract, and part of the evaluation of the national exemplars). This funding has become much more strategic, more directed and more closely connected to policy.

The Assessment Funding Pool is a separate appropriation, which goes through other Sections in the Ministry but also funds the work. The Comparative Education Research Unit contributes to international studies:

PISA (Programme for International Student Assessment)

TIMSS (Third International Mathematics and Science Study)

PIRLS (Progress in International Reading Literacy Study)

These studies inform the policy sections of the Ministry about New Zealand students compared with those of other countries and provide contextual information.

It is also common for new initiatives to have evaluation funding attached to the implementation funding. Important pieces of work may produce results that sit in isolation and apply only to a specific project or anything that it specifically applies to. Lack of such integration means that a project may not inform any longer-term vision. Examples are the evaluations of Technology and the Arts (Dance, Drama, Music and Visual Arts) curriculum initiatives. Until recently programme evaluation of this type informed the Ministry only in a piece-by-piece manner and the evaluation reports do not say anything about the wider picture of professional development. Current evaluation strategies in the Ministry of Education recognize and are working to redress this issue.

In the 2003/2004 year funding was also available through:

The Teaching and Learning Research Initiative (A budget that was announced in 2003 stated that funding will be distributed for field-initiated type work from the next financial year.) The Teaching and Learning Research Initiative was contracted to NZCER and directly funded from Vote Education

The Centres of Innovation where academics are going to be working with practitioners. Funded through Early Childhood Education

Building Capability in Special Education. There is a substantial programme funded through Group Special Education with the Research Division in an advisory role.

How does the Ministry of Education meet short, and long-term research needs based on both current policy needs and the provision of specific background information?

Student Support is an example of the Research Division developing a programme evaluation strategy to move away from the production of discrete studies that do not integrate to produce an overall picture. The Research Division had an opportunity to peer review the policy cabinet papers, in which the Policy Section said, "evaluations look good but never tell us what we need to know."

Research Division indicated to the policy sections that, without a *policy framework* that is explicit to the evaluator then it is very hard for the evaluation to meet policy needs. The Research Division is now involved in supporting the concept of a policy framework within which programme evaluation can be placed. The relationship between the Research Division and the policy sections, where the Student Support policy work was initiated (Group Special Education and Education Improvement and Support) is continuing to work towards development of a cohesive evaluation and monitoring framework for Student Support.

The Research Division's work in Early Childhood Education (ECE) parallels the points noted above. The 2001 budget shows, that instead of a separate line for evaluation in each of the initiatives in the Early Childhood Strategic Plan, there is a separate budget that is the Early Childhood Strategic Plan Evaluation. This helped fund a logic model diagram and draft programme logic matrix, developed by Patricia Rogers (2003) for the evaluation of the 10-year Strategic Plan. The model is similar to Funnell's (1997) matrix as described in Chapter Three. The "Student Support" framework is another strategy based on an approach that is closely connected to evidenced-based policy and a joint team approach.

While encouraging a more formative approach, as in the SEMO Intervention and the Numeracy Development Project where the implementation and the evaluations of the

programme initiatives are concurrent, the Research Division and Learning Policy Frameworks Team are aiming for a balance between process and outcome evaluations.

One informant suggested that evaluation has been a quiet revolution - like the fax machine - all of a sudden you cannot do without it. Initially the Research Division did not formally recognise the particular additional qualities required of evaluations in contrast to programmes of research. This recognition came with members' attendance at conferences and courses, and from reading national, and international research.

What processes are in place to ensure there is dialogue between the Ministry of Education and other Ministries that influence education policy development?

At a Ministry of Education management level there is an interdepartmental Senior Officials Group that represents Health, Education and Welfare. The Senior Manager of the Research Division is involved in an interagency committee of 20 government agencies with an interest in social policy evaluation and research known as SPEaR:

> The role of SPEaR is to oversee the Government's social policy research purchase. In particular, the:
>
> Improving co-ordination of social research and evaluation across agencies
>
> Improving the quality of social research and evaluation through capacity and capability building
>
> Focusing cross-sectoral priorities to encourage greater alignment of social research and evaluation with the Government's social policy priorities
>
> Providing communications hub and building connections in the social research sector
>
> Linking social policy research and evaluation evidence to policy decisions.
>
> (Ministry of Social Development, 2003, p. 3)

Close working relationships with the New Zealand Council for Educational Research (NZCER), the Prime Minister's Department, Cabinet and agencies such as the University Vice-Chancellor's Committee are being fostered. This cross-agency network is necessary for discussion and critique of policy. SPEaR chairman, Dr. Arthur Grimes (2003) states,

"poor information can lead to well-intentioned programmes failing to achieve their targeted outcomes or even making matters worse."

The synthesised approach to policy has attracted interest from within the Ministry of Education, and from other government departments. However, the difficulties should not be discounted. Departments have their own cultures and when working interdepartmentally different work cultures may have to be resolved in a common effort.

How does the Ministry of Education involve stakeholders in the programme evaluation and policy development process?

One of the issues in evaluation is how to come to terms with the complexity of the rights of stakeholders and the disbursement of power. Both have to be worked through. Beal (2003) points out that in reality a programme evaluation is contracted by one stakeholder, and the many other stakeholders who play a part in the evaluated programme may feel uneasy or put at risk by the action of the evaluation.

The Research Division views stakeholder input in programme evaluations as a sensitive issue. According to one informant, "in research you parcel out people, as if they are just a noise in politics. But in evaluation you ignore politics with a little p or a big P at your peril." It is therefore necessary to clarify the issues so that all stakeholders have realistic expectations of the process and the results.

How does the Ministry of Education involve Māori stakeholders in the process?

One of the greatest things to happen in the last five years according to Patton (1997) is the acknowledgement of cultural difference.

Stakeholders' input within Māori education is closely tied to Māori values. Many Māori actually have a group that they report to, their own stakeholders or whānau of interest. The issue within Māori education is where the control of education lies. The Research Division is developing a process that enables Māori researchers and the associated whānau to share control and have a level of input about the research needed and the method and the approach to be taken. It is about negotiation and it is about the kind of talking that seems to

make the process work, but it takes time, not only in research but also in the operational aspects of the projects.

What is programme evaluation's part in policy development?

An initiative, which could address the issues of inter-relationships, and evaluation informing policy, is the recent use in some Ministry of Education contracted evaluations of Rist's (1998) approach to policy making. Like Patton (1997) in his view of evaluation, Rist emphasises the view of policy making as a process. Rist sees policy making as "an ongoing set of adjustments and mid-course corrections that eliminates the bind of having to pinpoint the event – that is, the exact time, place, and manner – in which programme evaluation is influential on policy" (p.1003). He recommends that a policy cycle and qualitative research be used to translate policy intentions into policy and programmatic realities.

Rist's (1998) cycle has three phases:

> Phase one is the policy formulation phase where the policy issue at hand is clarified, what has previously taken place in response to this issue is reviewed and what is known from previous inquiry which may assist currently is identified.
>
> Phase two is the implementation phase and is concurrent with policy formulation and programme evaluation. During this phase the policy formulated is transformed into programmes, procedures and regulations. Programme evaluation addresses the issue of how to use available resources in the most efficient and effective way in order to have the most robust impact on the programme.
>
> Phase three is policy accountability. It is conducted when enough time has elapsed for questions of accountability, impacts and outcomes to be determined. In other words this phase assesses the consequences of the programme (p. 1003).

According to Hume and Bryce (2001), "we know from the literature on policy making that even where research evidence is clear – and often it is not – the various stages of policy development are often far from logical involving political intrigue, compromise, are subject to professional and bureaucratic self-interest, and beset with operational difficulties" (p. 343). However, the informants in this present study believe there is increasing evidence of programme evaluation effectively informing policy in the Ministry of Education.

It is acknowledged that the influence of evaluation on a programme is only one piece of the evaluation story but the Research Division believe they are increasingly producing evaluation reports that are robust enough to add to the general body of knowledge in order to inform future policy development. To assist in this, the Research Division is using techniques of meta-evaluation (evaluation of evaluations) to identify the common factors. This shifts evaluation to the research end of the spectrum.

When policy-makers receive the Research Division's programme evaluations there is still an element of "this is what it's showing - what does that mean for our decision making?" and they have to weigh that up in light of the current political situation and the finances available. As an informant described it: "We have to get the ducks in a row."

5.3 Policy development strategies favoured by the Ministry of Education

The government is requiring policy that is evidence-based. The evidence-based process is one that makes transparent the evidence that is connected to the issue in question, how that evidence is gathered, how it is analysed, and how that contributes to a conclusion. Within this the Research Division is not looking for a certain kind of knowledge but is looking for a process that is very well grounded in research methods for evidence-based policy.

Meanwhile, the Research Division acknowledges the importance of the policy view that evidenced-based policy means that the Minister can be told "well the weight of evidence suggests that this programme is quite expensive but we are getting great savings and having great impact in XYZ." The Learning Policy Frameworks Team believe that their members have the skills required to meet most of the evidence-based policy needs but they acknowledge that the input of researchers and programme evaluators is crucial in the process.

5.4 The New Zealand Government Analytic Frameworks

Information herein was gained from government documents and relevant literature as I sought to clarify and support the information gathered from my informants.

In pursuit of its focus on better learning for every New Zealander (Ministry of Education, 2003a) the government is encouraging the use of the analytic frameworks: (1) evidence-based policy, (2) best evidence synthesis, and (3) intervention logic.

Hammersley (2002) reminds us that it is only recently that the term "evidence-based" has appeared in education, a field that Scriven (1983) describes as being like medicine in that its name simultaneously refers to a practice and to a field of disciplinary inquiry. Evidence-based policy was initially conceptualised within medicine and is now commonly incorporated within the field of education. While the focus in medicine has been on the quality of practice, in education it has been on the quality of research (Hammersley, 2002). It tends to support quantitative and statistical methods.

Rather than supplying or validating effective techniques or policies, the benefit is now believed to lie more in raising questions about current assumptions, and supplying alternative perspectives on the work of teachers, education managers and policy-makers, and on the contexts in which they operate i.e. programme evaluation. Evidenced-based policy is an approach, which argues that policy and practice should be capable of being justified in terms of sound evidence about their likely outcomes, but in view of the differences discussed earlier, sound evidence depends on one's interpretation.

Pawson (2001b) feels programmes are:

> over-saturated with theories. Policy architects seek the big ideas whose time is right, practitioners work with folk theories gained from workday victories, subjects have theories of "self" justifying their own behaviour, and occasionally some orthodox social science theory finds its way into the construction of initiatives. (p. 7)

Pawson believes what is needed in evidence-based policy development is:

> quantitative research to gauge effect; qualitative research to gauge stakeholder response; documentary work to provide contextual information; "textual" and "semiotic" analysis to show how information may be interpreted; and the production of comparative research to show the shaping forces of locality and time. (p. 18)

The only methodological standard that counts is that the evidence should bear the weight of the particular claim made for it.

There is nothing new about the idea that the best available evidence should inform policy. However, Smith (2002) lists seven "enemies" of evidence-based policy as seen by the Public Service employees as: bureaucratic logic; the bottom line; the difficulty of achieving consensus; politics; intransigent culture; cynics; and time. Smith also lists seven "enemies" of evidence-based policy as seen by researchers as: the scientific model of research; disciplinary purity, problems which are "too complex to address"; professional imperatives such as scientific publication; exposure to policy debate and political arena; uncertain funding; and professional disdain.

According to Pawson (2001b) knowledge should "speak to power and its voice should be the voice of the researcher. However, knowledge speaks in a whisper, as empirical inquiry simply cannot make its voice heard amidst the clatter of other political imperatives in policy-making" (p. 19). This accords with the opinion of personnel of the Ministry of Education's Research Division who state that they are but one small piece of the policy-making process.

Hume and Bryce (2001) believe that with the increasing emphasis on evidenced-based policy initiatives, there is a growing tendency in public debate for "knowledge," "research" and "evidence" to be used interchangeably (p. 330). However, evidence is always inextricably intertwined with the actions, interactions and relationships of practice. Lau (1997) contends, "if the evaluation design is properly structured and relationships are appropriately defined, the…evaluators can provide an important safeguard, ensuring requisite quality of evidence validity and credibility of the evaluation undertaking" (p. 13).

Best-Evidence Synthesis

The words of a six-year old Swedish boy are relevant to the provision of policy-makers with the best possible evidence:

> If I was prime minister, I would decide good things. But I don't really know what good is, so I suppose I'd have to guess a little. (Annemalm & Nilsson, 2003)

Slavin (1987) says best evidence synthesis can only occur when criteria are identified for what is good quality research, which yields best evidence in a particular field. This perspective provides the benchmark for evidence-based policy. The central idea underlying best-evidence synthesis is that the most useful data to support policy decisions will be distilled from a synthesis of all existing research in that particular policy domain. Pawson (2001a) says this is to provide a bridge between knowing and doing, but it brings up the issue of how to undertake the synthesis. Traditional approaches are meta-analysis and narrative review.

By 2004 the Ministry of Education had commissioned five best evidence syntheses iterative:

> *Quality Teaching - Early Foundations*
>
> Farquhar, S. (2003). Childforum Research Network
>
> *Quality Teaching for Diverse Students in Schooling*
>
> Alton-Lee, A. (2003). Medium-Term Strategic Policy Section. Ministry of Education
>
> *Characteristics of Professional Development Linked to Enhanced Pedagogy and Children's Learning in Early Childhood Setting*
>
> Mitchell, L. & Cubey, P. (2003). NZCER
>
> *The Complexity of Community and Family Influences Children's Achievement in New Zealand*
>
> Biddulph, F., Biddulph, J. and Biddulph, C. (2003). Biddulph Education Consultants.

All are linked to the outcomes of diverse students and are based on the government's goals (refer p. 27 of this thesis). The purposes of the syntheses are:

> To systematically identify and bring together, evaluate, analyse, synthesise, and make accessible, relevant evidence linked to a range of learner outcomes,
>
> To bring a systematic approach to evaluating what research reveals works, in order to illuminate policies, contexts, systems, resources, approaches, practices, alignments and influences impact on diverse learners,
>
> To deepen understanding of what works in education,
>
> To strengthen education policy and educational development in ways that address patterns of systematic under achievement,
>
> To build knowledge that will strengthen systematic responsiveness to, and educational outcomes for, Māori,
>
> To build knowledge that will strengthen systematic responsiveness to, and educational outcomes for, Pasifika students (p. 4, Alton-Lee, 2004).

Synthesis should be theory driven and should treat each intervention as a case study. Programmes are not portable but ideas are. The goal of synthesis should be transferable theory and transformations of information suited to the circumstance. According to Mark, Henry and Julnes (2000), the goal of programme evaluation and synthesis is that of "making sense" (to assist, support and extend natural human abilities) to observe, understand and make judgments about policies and programmes.

Best-Evidence Use

According to Nutely et al., (2002b) policy-makers the world over report very little targeted and deliberate use of programme evaluation findings. They believe attention is more likely to be paid to evaluation findings when the evaluation is timely; the evidence is clear and relevant; and the methodology is relatively uncontested. Policy-makers believe in evidence as an important counterbalance to expert opinion.

Programme Logic

Clear and logically consistent methods have not been readily available to help programme managers make implicit understandings explicit (McLaughlin & Jordan, 1999). To this end, programme logic is a framework through which to carry out intervention logic. As previously stated, the New Zealand Government's preferred framework for programme logic is that of Funnell (1997). The Research Division has also worked extensively with Patricia Rogers. There are simpler frameworks available but they do not identify factors within and outside the control of management. It is based on the theory of the causal links between resources and activities, outputs, short-term impacts and long-term outcomes. When speaking to Wellington College of Education staff on the Te Kauhua: Māori in the Mainstream Pilot project evaluation in 2003, Higgins (personal communication) stated programme logic is needed as a form of support for critical analysis of interventions and for creating common frames of reference that are especially helpful when programmes involving Māori/Pasifika are being evaluated.

In the SEMO evaluation simplified programme logic was used to identify contextual factors, strategies, intermediate goals and long-term outcomes (Robinson, Timperley & Bullard, 2000).

An example of Funnell's (1997) programme logic matrix appears on page 58.

Table II **Adapted from the Programme Logic Matrix Structure (Funnell, 1997)**

1. Outcomes Hierarchy e.g. Changes in knowledge and skills of the target group	2. Success Criteria e.g. What are the desired types of clients	3. Factors Within the Control of the Programme e.g. Quality service delivery	4. Factors Outside the Control of the Programme e.g. The demographics of the target group.	5. Programme Activities and Resources e.g. Training given to staff to improve service quality	6. Performance Information e.g. The percentage of clients who show improved knowledge	7. Comparisons (To judge and interpret performance data). e.g. Comparisons with standards to make judgements
Ultimate Outcome						
Intermediate outcomes						
Immediate Impact						
Outputs						

Programme Logic is a strategy, which can be used to evaluate government activities, to improve the reporting of results to Government, and for informing policy. An example of its use within the Ministry of Education occurs in the *Evaluation of Two ECD Services* (Mitchell & Mara, 2001). However, it is important to remember that it is but one input in evidence-based policy.

5.5 Summary

In 1996 the appointment of eight new division managers within the Ministry of Education encouraged a more direct connection between policy and research. As a consequence, the bidding process for research funding has changed and longer-term formative programme evaluations are being undertaken. The three-year political cycle still presents an issue in programme evaluation timelines as the politicians often require information before a quality programme evaluation can be sufficiently validated.

As the Ministry of Education has reviewed its processes and practices, issues have arisen in the Research Division's capacity to meet the needs of government and the education community. Highlighted in this is the need for the Research Division to increase its relationship with the policy sections. Many of the programme evaluations which have been carried out sit outside the Research Division and their control are with the various policy managers who initiated them. This increases the need for interaction.

In the pursuit of effective outcome-based management the New Zealand Government is requiring evidence-based information and encouraging the use of analytic framework, such as Intervention Logic.

The use of Rist's (1998) approach to policy making is a new influence seemingly with potential in the Ministry of Education applications. Within this approach a policy cycle is used where qualitative research informs each of three phases (formulation, implementation and accountability) to translate policy intentions into policy and realities.

Chapter Six gives an overview of the complex SEMO programme evaluation in which the evaluators "designed the evaluations to produce both qualitative and quantitative data" (Robinson, Timperley, Ward, Tu'uga Stevenson, & Mitchell, 2004, p. x).

CHAPTER SIX

6.0 Strengthening Education in Mangere and Otara

6.1 Description

This chapter outlines one of the two programme evaluations selected to illustrate the developing relationship between programme evaluation, research and policy within the New Zealand Ministry of Education. My informants from the Research and Evaluation Team recommended the SEMO (Strengthening Education in Mangere and Otara, 1998-2004) evaluation as an example of good practice in programme evaluation in the formative process of policy development.

Reports have been written by the evaluators (Timperley, Robinson and Bullard, 1999; Robinson, Timperley and Bullard, 2000; Phillips, McNaughton and MacDonald, 2002; Timperley and Lam, 2002; Timperley, Phillips & Wiseman, 2003, Robinson and Timperley with Ward, Tuioti, Tu'uga Stevenson and Mitchell, 2004) some of whom were also involved in the programme development and its delivery. ERO has also reported from a more general perspective on education in Mangere/Otara (Education Review Office, 1996, 1998).

The SEMO intervention is one of several long-term, area-wide interventions being implemented by the Ministry of Education. Robinson et al. (2004) state the "intention was to increase the capacity of the schools and communities of Mangere/Otara to offer high quality learning environments for children" (p. ix). To begin, the Ministry provided additional assistance to individual or clusters of schools with the aim of strengthening the schools' capacity to be self-managing. The National School Support Section, Ministry of Education, coordinates such interventions. At any given time approximately 15% of the nation's schools are assisted through this policy (Robinson et al., 2000).

6.2 The Problem

Mangere/Otara is a South Auckland urban area with a predominantly Māori and Pasifika population. The area had seventeen decile 1[4] schools within it: two high schools (Years 9-

[4] Decile 1 - the lowest socio-economic status allocated from the Census statistics

13); one middle school (Years 7-10); two intermediate schools (Years 7-8) and 12 primary schools (Years 1-6).

Educational issues related to low education standards appeared in the 1981 report, *Tomorrow May be Too Late,* which highlighted the disadvantage suffered by most school age students in the Mangere/Otara area. The report is commonly referred to as The Ramsay Report, taking the name of the chairman of the team responsible for its preparation. In addition, ERO (Education Review Office, 1996) identified 42% of schools in the Mangere/Otara area as performing very poorly, 27% of the schools at the greatest risk of non-performance, and 15% of schools under-performing. ERO stated that there was no sustainable high quality education being provided in Mangere/Otara schools and that the Boards of Trustees, school management, and teaching needed to be improved. The review of the Mangere/Otara area was the first time ERO had investigated the overall performance of a group of schools. They subsequently reviewed other low socio-economic demographic areas in the East Coast, 1997, the Far North, 1997, and the West Coast, 2003, and identified many similar issues.

The New Zealand Education Act, 1989 and State Sector Act, 1988 precluded direct state intervention to governance and management matters beyond what schools agreed to; that is, unless the schools were violating statutory regulations or threatening the viability of the school.

The Education Act 1989 did not provide any incentive for self-managing schools to adopt a collaborative approach to educating their students. The self-governance model established school autonomy but required schools to take no responsibility for any inter-school implications resulting from their operations (Robinson et al., 2000).

The Mangere/Otara schools were experiencing escalating professional and inter-school conflict as a result of falling rolls. Many families by-passed their local schools to seek a perceived "better education" in other areas. Schools in other areas introduced new bus routes to pick up these students. The local schools saw this as poaching and made efforts to market their schools by holding public recruitment drives. This led to a loss of confidence by the community in the Mangere and Otara teachers and/or schools resulting in up to 40% transience of students in some schools and a high staff turnover (Robinson et al., 2000).

In 1997 the Ministry of Education described the intermediate and secondary education in the Otara area as showing a greater decrease in roll than anywhere else in New Zealand. Up to 70% of secondary students in the geographic area were enrolled in schools outside the area. Otara was in danger of losing a secondary school (Robinson et al., 2000). From the government's point of view this meant that inefficient use was being made of Crown assets.

6.3 Ministry of Education Response

In response to the Education Review Office Report (1996) the government funded initiatives totalling $22 million over three years in the Mangere/Otara area to address the educational problems (Education Review Office, 1998).

The SEMO intervention implemented programme initiatives through five phases and it continues to the present time. Many of the programme initiatives overlapped and are ongoing with some resulting from the findings of the evaluation of the programme's prior phases. The phases are described in Figure 4.

The intervention was complex. During the first six months of the SEMO initiatives, a SEMO Team was established to spearhead the projects. The Ministry regarded it as vital that there was an ethnic mix on the SEMO Team (1 co-ordinator, 1 female Māori facilitator and 3 female Pasifika team members). Together with personnel from the Ministry, the team focused on consulting with stakeholders about their perceptions of education problems and priorities.

The programme initiatives that emerged included:

The Structure of Schools in Otara

Three-way Partnership Involving Aspects of Literacy

Boards of trustees as partners

The Early Childhood Primary Link (ECPL)

Māori Education in Mangere/Otara

The intervention and the programme evaluation took place at the same time. There was also an overlap between those who carried out the evaluations.

Phase One – June 1996 - October 1997

The Education Review Office (ERO) reports, research and consultation resulted in the Ministry presenting a paper *Schooling in Mangere and Otara: A Strategic Improvement*, to the then Minister of Education, the Hon. Wyatt Creech.

Phase Two – November 1997 - April 1998

Evaluation activities began in 1998 to the extent to which the SEMO intervention focused on, and improved the capacity of schools to monitor and improve the quality of the learning environments they provide for their children.

Phase Three – May 1998 - December 1998

A continuation of support for the first SEMO project, *Communities in Schools via Literacy* (CISvL) that the Crown funded through the FPA (Funding Provision Agreement). Evaluators prepared a scoping report following which a co-ordinator and team were appointed to identify and analyse the needs of the 45 schools in the area. Literacy teaching and learning was identified as the context in which to judge the capacity of the schools to monitor, analyse and improve student achievement levels.

Phase Four – January 1999 - December 2000

The CISvL projects continued to be developed and implemented in the schools and were monitored by both ERO and the Ministry of Education. A second project, Early Childhood Primary Link (ECPL), looking at the transition from early childhood centre to school was implemented. Evaluators prepared a scoping report following which a co-ordinator and team were appointed to identify and analyse the needs of the 45 schools in the area. Literacy teaching and learning was identified as the context in which to judge the capacity of the schools to monitor, analyse and improve student achievement levels.

Phase Five – January 2001 -

A review of the structure of schooling for Otara and the Middle School option proposed by the two intermediate schools was implemented.

The role of the Board of Trustees as partners and employers, and the expertise they brought to the job was examined. Professional development - Analysis and Use of Student Achievement Data (AUSAD) was implemented as a direct result of the previous oral and written programme evaluation reports to the Ministry. Studies of the three-way partnership between the Ministry, schools and their communities continue.

Figure 4. Phases of the SEMO Intervention Model

6.4 The Programme Evaluation

The evaluation of the SEMO intervention was contracted to the Auckland UniServices Ltd, a wholly owned company of the University of Auckland, at $507,785 for phases 1, 2 and 3. The evaluators supplied a number of interim reports on specific aspects of the intervention before the 2004 final report. These included a scoping report in 1998 (Timperley et al., 1999) that discussed how to improve the quality of relationships through partnership between schools, communities and the Ministry of Education. Recent research cited in Robinson et al. (2000) on school improvement indicated that sustained gains were more likely if interventions involved both schools and community groups. The rationale was that comprehensive interventions were more likely to identify and align the resources of communities and professionals in the pursuit of student learning.

However, any partnership between the Mangere/Otara community and the Ministry was going to be difficult. Timperley et al. (1999) reported that a long line of Ministry of Education representatives (mainly Pakeha) had come and tried to tell the community how to run schools in the area and left with no positive outcome and, as mentioned before, schools were expected to be self-managing.

Within the programme evaluation the evaluators sought answers to two questions:

> To what extent has the partnership between schools, early childhood centres, the wider community, and the Ministry of Education been developed?
>
> Do these partnerships deliver sustainable improvement in achievement for the students of Mangere and Otara? (Robinson et al., 2000)

The evaluators chose a formative approach to the programme evaluation and used Robinson and Walker's Problem-Based Methodology (PBM) (cited in Timperley et al., 1999) to address the risk that while preserving the relationship through which the formative influence was achieved the evaluators could lose critical independence.

The evaluators describe this particular methodology as treating all policies and practices as solutions to problems, however unsatisfactory those solutions might be. In using PBM the

evaluators went beyond Scriven's definition of evaluation. Not only did they judge merit through formative evaluation (evaluation designed and used to improve an object, especially when it is still being developed) but they also sought to explain why aspects of the intervention operated as they did. According to Timperley et al. (2003), the evaluation was a "process" evaluation. It can be seen that the evaluation proceeded against a changing background of results.

The evaluators took into account recent national and international research on school improvement, constraints arising from national education policy, and the culture, politics and recent history of Mangere/Otara (Robinson et al., 2000). They developed a framework that was intended to be responsive to, but independent of the programme itself. Stakeholders were asked for their perceptions of the students' needs and priorities. The child-related needs chosen were self-esteem, literacy and exposure to learning environments (Timperley et al., 1999). To assist the intervention in meeting the child-related needs, the authors identified school and community needs of sustainability and educative partnerships between school, community and the Ministry of Education.

Part of the evaluation methodology was to seek feedback on each of the programme initiatives from key stakeholders and invite their critical comment. In their reports to the Ministry, the evaluators refer to the collaborative relationship developed between themselves and the local and national School Support Officers of the Ministry of Education. "Four or five Ministry of Education officials with responsibility for policy and operational aspects of school governance from both regional and national offices, had responsibility for the SEMO initiative" (Robinson & Timperley et al., 2004, p. 4). The evaluators' reports claim that their findings have influenced decision-making and much of recent policy developed by the Ministry appears to address issues that were raised in the SEMO programme initiative evaluation results, and which have expanded to national initiatives. For example, the National Assessment Tools[5] were developed to assist teachers in analysing student achievement data.

---

5 National Curriculum Exemplars, asTTle

The Structure of Schools in Otara

Resolution of the problems in Otara of school numbers, and the issues resulting from applications by the two intermediate schools for middle schools status only succeeded when a Forum of Chairpersons of all participating Boards of Trustees was established and two professional facilitators were employed. A Ministry representative from the National School Support Sector attended all 20 meetings of the Forum but principals were excluded in phase two, as they had tended to dominate group thinking in phase one.

The Three-way Partnership Involving Aspects of Literacy

In phase one when schools had the freedom and resources to develop projects they saw as important, successful projects resulted but they were generally school management based and had little impact on the overarching purpose of the intervention, the raising of student achievement.

In phase two the focus of school projects moved to the core functions of teaching and learning, and recognition of the importance of annual school self-review. The information collected resulted in the Communities in Schools via Literacy (CISvL) initiative designed to involve parents and community in children's education.

As a baseline measure by which to judge the CISvL initiative the evaluators collected data on student achievement in reading, writing and spelling from 26 schools and carried out interviews with literacy leaders and teachers.

Their findings were:

Assessment data were not being used to inform programmes

The majority of literacy leaders could not interpret the data

Literacy leaders seldom supported their judgement of achievement with data

Most schools reported student achievement data to parents in the form of ratings that were not understood by them.

The issue of teacher assessment capacity led to the implementation of the Analysis and Use of Assessment Data (AUSAD) initiative in the SEMO area. Robinson and Timperley et al. (2004) believe this to be "one of the most important initiatives in the SEMO project" (p. ix). It is not always easy to see what came first but it would appear that the National

Literacy Leadership professional development, National Assessment Tools Development, the Revised National Administration Guidelines (2000) in which literacy and numeracy became a priority for all students Years 0 – 4, and the increase in the power of the Minister to intervene in a school where deemed necessary, all address issues raised in the SEMO programme evaluations around teacher capacity in the area of student achievement data analysis.

Boards of Trustees as Partners

The Education Review Office (1996) had reported that trustees did not understand their role and neglected their responsibilities. However, the evaluators felt that the skill level of trustees to carry out the task was a larger issue and expectations of them were not realistic.

Training for new trustees has been available since 1996/7 with increased training offered after each trustee election. In 1998 training was offered to new trustees and Boards of Trustees governing schools seen as being "at risk". Because of the lack of people with the necessary skills and the demanding nature of the trustee position, schools are now able to put half their trustee positions up for election eighteen months after national elections (the half way point of the election three year cycle). This not only ensures greater retention of skills, knowledge and stability on the Board, but is an opportunity to seek those with the required skills.

Early Childhood Primary Link (ECPL)

To address the issue of transition between a variety of early childhood centres and primary schools a co-ordinator was appointed to develop continuity. It was found that there was a lot of interaction between schools and early childhood centres at a relationship level but not at a working level. There was a lack of understanding of each other's purpose and expectations.

Picking up the Pace

As a component of the ECPL initiative, the Gwenneth Phillips professional development course "Picking up the Pace" was implemented. This catered for teachers of five-and six-year-olds and the literacy leaders from seven SEMO schools. Student achievement information was collected through running records (text reading) and the Burt test (word recognition). The evaluators (Phillips, McNaughton & MacDonald, 2002) claimed that the programme initiative indicated that sustained achievement of students in decile 1 schools was possible. They reported that contextual factors were not significant and that once teachers acknowledged this they raised their expectations of both the students and themselves.

The reports have been criticised by some researchers, i.e. Nash (2004), and the Education Policy Group (2003) at Massey University, and this topic will be addressed further in the next chapter. Despite this, Ministry of Education policy has implemented the Picking up the Pace programme in schools in socio-economic areas, such as, Porirua East and Wainuiomata.

Māori Education in Mangere/Otara

The proposed evaluation of the contribution the SEMO intervention made to the education of Māori children in Mangere/Otara did not take place. A Māori group, Te Ropu Whakangungu Matauranga Taiohi mo Apopo, preferred to commission their own research and so the evaluation did not proceed (Robinson et al., 2000).

However, the SEMO evaluators recommended two studies related to Māori education in Mangere/Otara be carried out:

> Developing Partnership between Māori and the Ministry
> Māori Parental Aspirations and Participation.

The Ministry of Education's Strategic Intent 2003-2008 (Ministry of Education, 2003a) indicates the government's commitment to improving education for Māori through the development of iwi partnerships.

Initiative Achievements

The evaluators identify the three main achievements of SEMO as:

1. Drawing attention to, and focusing resources on, the task of sustainable improvement in student achievement
2. The development of relationships through which the task of sustainable improvement can be accomplished
3. Having developed a culture of evolution and learning able to detect and correct mistakes, identify and avert political risks and respond to the experiences of those whose interests the initiative are designed to serve. (Robinson et al., 2000)

6.5 Summary

The Ramsay Report (1981) and ERO (1996, 1998) reports had identified issues of low-achieving students in the Mangere/Otara area for many years. The problem was compounded by the New Zealand Education Act (1989) and State Sector Act (1988), which provided no incentive for self-managing schools to adopt a collaborative approach to problems and restricted the ability of the Ministry of Education to intervene in the running of the schools.

In response to the adverse reports, the Ministry of Education funded programme initiatives within the SEMO intervention. The initiatives in Otara covered issues of school structure, school governance, and the analysis and reporting of student achievement data. The intervention worked through a series of phases with many of the programme initiatives overlapping. All programme initiatives are ongoing – with some resulting from the evaluators' findings.

Chapter Seven presents an overview of a very different Ministry of Education programme evaluation. While the SEMO intervention is complex, and restricted to one demographic area, the Numeracy Development Project is a national initiative that concentrates on developing teacher capability in the teaching and learning of number.

CHAPTER SEVEN

7.0 Exploring Issues in Mathematics Education

7.1 Description

Discussions similar to those I held with the Research and Evaluation Team were held with personnel from the Learning Policy Frameworks Team. As an example of an interdepartmental, formative programme evaluation process, the section recommended the Numeracy Development Project: Exploring Issues in Mathematics Education (the Numeracy Development Project). The information on the programme was gathered from reports by the evaluators (Higgins, 2001, 2002, 2003; Higgins, Parsons & Hyland, 2003; Irwin, 2003; Irwin and Niederer, 2002; Thomas and Ward, 2001, 2002; Thomas, Tagg and Ward, 2003).

7.2 The Problem

Concern over achievement of New Zealand students in mathematics was highlighted in 1981 in the Second International Mathematics Study (SIMS) (Visser, 1999). Thirteen years later the Third International Mathematics and Science Study (TIMSS) (Visser, 1999) found that the overall levels of mathematics achievement of students Years 8 and 9 were slightly lower than in most other western countries. It also noted that Māori and Pasifika students were achieving below Asian and Pakeha students. However, as in schools today, there were high and low-scoring students within each ethnic group.

In 1992 the mathematics community and the Education Review Office (1994) signaled difficulties in implementing the 1992 *Mathematics in the New Zealand Curriculum.* Deficiencies in students' performance were also highlighted in the National Monitoring Project (NEMP) (Flockton & Crooks, 1997).

7.3 Ministry of Education Response

The Research and International Section of the Ministry of Education commissioned a further report on mathematics by Holton, Spicer, Thomas, and Young (1996). Continued investigation was considered necessary and in 1997 the Ministry of Education set up a Mathematics and Science Taskforce.

In response to the needs identified by the Taskforce, the Research and Curriculum Divisions of the Ministry of Education commissioned literature reviews on the issues. A Research Seminar on Mathematics Education (Year 0-6 Students) was held on 12 June 1998 where the researchers were also asked to build a conceptual framework for the development of a mathematics education research programme.

Following this process the Ministry of Education developed the following Key Strategic Outcomes (Higgins, 2001): improved student achievement; improved teacher capacity.

The involvement of a number of groups in the mathematics community was sought and the Numeracy Development Reference Group was formed. This made it possible for policy to be based on available research information (Higgins, Parsons & Hyland, 2003). In pursuit of their strategic outcomes the Ministry of Education implemented an intervention programme, the Numeracy Development Project. The aim of this project was to better understand students' number strategies, knowledge, and their stages of development. It was expected that improved teacher knowledge would lead to improved student achievement.

The Ministry of Education began the Numeracy Development Project with a pilot study of the Australian programme called Count Me in Too (CMIT). This programme formed the first phase of the intervention and was delivered through a professional development programme in which teachers were expected to reflect critically on their own mathematical pedagogy and content knowledge (Thomas & Ward, 2001). The professional development model was school-based with numeracy facilitators working alongside classroom teachers modeling teaching strategies and observing teachers implementing the strategies.

The desire to ensure continuity across Years 1-8 led to the trialing of an extended Number Framework as part of the Numeracy Project implementation process (Young-Loveridge & Wright, 2002).

In 2000, the National Education Guidelines (Ministry of Education, 2000) were reviewed and numeracy was made a priority for all students Years 1-4. This led the Numeracy Development Reference Group to define numeracy and make strategic links between the Numeracy Development Project and the New Zealand national mathematics curriculum (Ministry of Education, 1992).

If teachers are to be effective in raising student achievement in Mathematics they need to be familiar with the strategies and knowledge contained in *The Number Framework* (part of which is in Appendix Four of this thesis). The strategies focus on how students solve number problems and the extent to which they use mental processes as part of the solution. It has a knowledge component that is broken into four content domains and a written recording component.

The Numeracy Development Project Assessment (NumPa), consisting of the Diagnostic Interview, is a tool that takes the form of an individual interview with students and reveals mental strategies used by students in solving number problems. Once the interviews have been completed teachers group the students according to the strategy stages aligned with each operational domain. They then use activities to support the development of both the strategy and the knowledge components.

7.4 Intervention Phases

The imported Australian programme, CMIT (Count Me in Too), was found to need adapting to increase its suitability for New Zealand conditions (Higgins et al., 2003). In 2000, a Years 4-6 exploratory study was implemented.

CMIT, which became The Early Numeracy Project (ENP) Years 1-3, was the first phase of the intervention. The Counting On programme that became The Advanced Numeracy Project (ANP) was implemented at Years 4-6, and The Numeracy Exploratory Study (NEST) focused on Years 7-10, followed. In 2002 Te Poutama Tau was implemented in Māori immersion units and in 2003 the Intermediate Numeracy Project began.

The phases implemented 2000-2001 are illustrated below:

Count Me in Too Pilot (2000) Years 1- 3

- Across all regions of New Zealand
- Data collected from all participants (17 facilitators, 563 teachers, 9309 students) through questionnaires and interviews, and the diagnostic test results

Year 4-6 Exploratory Study (2000 Terms 2-4)

- A small study in Auckland, Waikato, and Wellington with emphasis on lower decile schools.
- Data collected from 12 teachers through concept-maps and semi-structured interviews.
- One focus was on the question 'what knowledge do you need to teach Mathematics effectively?'

Advanced Numeracy Project (ANP) (2001)

- Formerly the Year 4-6 Exploratory Study
- Data collected from all 40,000 students through the ANP assessment tool and 480 teachers, 17 facilitators and 70 principals through questionnaires
- Case Studies of 4 facilitators, 8 teachers and their principals through observations of facilitators, reflections and interviews

Numeracy Project Exploratory Study Years 7-10 (NEST) (2001)

- Small study in 12 secondary schools and 6 intermediate schools in six centres
- Data collected from comparative assessments of 4000 students on the NEST assessment tool. Interviews with teachers and facilitators.

Figure 5. Phases of The Numeracy Development Project
(Based on information drawn from evaluations of the Numeracy Development Project)

Although it is a long way off, the implementation of these phases could see successful strategies implemented at all levels of student education from Year 1 to NCEA (National Curriculum Education Assessment). Greater consistency in teacher knowledge, and a consequent raising of student achievement could result.

7.5 Programme Evaluation

At the same time as it was implemented, a programme evaluation of the Numeracy Development Project intervention was contracted by the Learning Policy Frameworks Team of the Ministry of Education. The evaluation of the contributing school (Years 0-6) section of the programme was contracted to evaluators from the Dunedin and Wellington

Colleges of Education. The Exploratory Study into Year 7-10 Mathematics was contracted to Auckland UniServices Ltd., wholly owned by the University of Auckland.

According to the evaluators, the Project is focused on the classroom, learner dynamics, content knowledge, assessment and pedagogy. Teachers use the assessment interview to gather, analyse and use data. Facilitators model good practice, observe teachers, provide feedback, and encourage reflection (Higgins et al., 2003). An important factor is the involvement and commitment of the school leadership. This commitment of leadership is stipulated in a Memorandum of Agreement signed by the Principal and the facilitator.

The evaluators sought answers to the following broad questions:

What knowledge does a teacher need to teach mathematics effectively?

Is change in professional knowledge reflected in classroom practice?

What influenced the change?

Does the programme impact on facilitator professional knowledge? And most importantly,

What shifts have there been in student achievement?

However, the programme evaluation includes many other additional questions.

To answer the main questions the evaluators used questionnaires, carried out case studies, used data from student diagnostic interviews, held interviews with teachers and facilitators, and undertook a longitudinal study of a group of teachers who had been in the development from the outset, to gauge the sustainability of the project.

7.6 Policy Development within the Numeracy Development Project

The formative nature of the programme evaluation of the Numeracy Development Project illustrates policy as an iterative, developmental process involving the mathematics community. The initial emphasis for policy development was on teachers' professional capability. Information to develop policy was drawn from existing evidence, followed by analysis of previous attempts at solving the problem of low teacher capability and how this information could be used in the current formulation of policy (Higgins et al., 2003).

The evaluators used Rist's (1998) framing of the policy process to analyse the Numeracy Development Project. This framing describes a dynamic approach to the policy process through coordination of aspects of formulation, implementation, and evaluation. Rist (1998) redefined decision-making in policy creation as "an ongoing set of adjustments, or mid-course corrections that eliminates the bind of having to pinpoint the exact time, place and manner – in which research has been influential over policy" (p. 1003). Rist (1998) also commented on ways in which qualitative research can be influential in each phase of the policy process because it is "longitudinal by nature, done in naturalistic settings, and focused on the constructions of meaning developed by the participants, it is in a unique position from which to assess the possibility of tools having the impacts intended by policy-makers" (p. 1014). This approach was also used in the evaluation of Te Kauhua undertaken by Tuuta, Bradnam, Hynds, & Higgins with Broughton (2002) who stated that the approach is new to the government and is distinctly better for Māori.

The Learning Policy Frameworks Team believes the Numeracy Development Project evaluation design is an excellent example of policy integration, feeding the evaluation results back and leading to changes. This dynamic and coherent process provides opportunities for stakeholders to raise issues and these become the next step in the policy development process. This exemplifies Patton's utilisation-focused approach to programme evaluation.

The Ministry of Education 2001-2002 Annual Report (Ministry of Education, 2002) stated that the National Education Monitoring Project (NEMP, 1997) showed that the Numeracy Strategy, of which the Numeracy Development Project is part, had resulted in substantial overall increase in Year 4 students' achievement.

Initiative Achievements

According to the reports to the Ministry of Education the implementation of the Numeracy Development Project has been a success on many levels. The *Number Framework* gives teachers directions for responding effectively to students' learning needs. The diagnostic interview focuses on student mathematical thinking (Higgins et al., 2003). The professional development took place in the teacher's classroom and the facilitators were able to make links with the intended programme and help the teachers modify their pedagogical practice.

Higgins et al., (2003) contend the learning community that developed around the project had high levels of trust, confidence and ownership. Leadership of the project was shared so that those involved, including Ministry personnel, worked to their strengths. There was provision for meetings at regional and national levels to foster open communication. Contribution was valued and new knowledge was being generated at each component in the project.

Nutely et al., (2002a) agree that research suggests that facilitation may be a key variable in the implementation of programmes. The evaluators found that the facilitators played a key role in mediating the policy for teachers. The components of facilitators' knowledge of mathematics, pedagogical knowledge, and their knowledge of learners' cognition in mathematics were significant in their success in delivery of the Numeracy Development Project. The facilitator quality brought about shifts in what happened in classrooms. In an area where Māori and Pasifika students did better than other students a small case study was set up to investigate the facilitator's role in this.

The teachers responded positively to reflective practice emphasised within the professional development.

7.7 Summary

Concern over the achievement of New Zealand students in mathematics was highlighted in 1981 and 1999 by international research studies. The responsibility for this situation was widely attributed to lack of quality teaching.

Teacher difficulties in trying to meet the requirements of the New Zealand National Mathematics Curriculum (1992) were identified by many within and outside the

mathematics community. In response the Ministry of Education set up the Mathematics and Science Taskforce, commissioned literature reviews, and held a Research Seminar on Mathematics Education for Year 0-6 students.

To address the issue of underachievement the Ministry of Education began the Numeracy Development Project in 1999 with the Count Me in Too project. A key element within this project is the Numeracy Development Project Assessment, a diagnostic tool designed to provide accurate information about the strategies and knowledge of students.

The Numeracy Development Project implementation and its evaluation were concurrent and incorporated evidence-based professional development which the Ministry of Education believe is the key element in improving teacher quality and raising student achievement. Within this formative policy development is an iterative, developmental process involving the mathematics community.

The initial emphasis for policy development was on teachers' professional capability. The worth of the policy was evaluated at each stage of its development with appropriate policy adjustments being made at each phase using Rist's (1998) framing of the policy process. The findings of the programme evaluation have been significant in shaping policy. The implementation has been reported as a success on many levels.

Chapter Eight brings all the information together in a discussion on programme evaluation, its uses, and standing in New Zealand today.

CHAPTER EIGHT

8.0 Summary and Discussion

In preparing this thesis I began with the idea of carrying out a case study of some programme evaluations done by the New Zealand Ministry of Education or under contract to the Ministry. This original intention expanded into an exploration of the origin of programme evaluations. Where did they come from? Then further expansion into why the Ministry was interested in programme evaluation. In turn, this led to consideration of the link between evaluation and policy, the institutional arrangements within the Ministry of Education which led to the purchase of evaluations of programmes, the style of evaluation supported by the Ministry, and provision for the use of the findings in determining policy.

Along the way I found I needed to explore whether there was a distinction, if any, between research and evaluation. I returned to my original intention of studying reports of some programme evaluations and, on the recommendation of officers of the Ministry I chose the SEMO and Numeracy projects looking not only at what they were but also for their outcomes in terms of shaping policy.

In exploring the disparate aspects of programme evaluation I read about the types of programme evaluation, and some of the methods and issues relating to it including the fraught topic of the relationship between research and evaluation and their respective status. I sought information about the origins of programme evaluation and about its use in New Zealand as an aid in developing policy in education. I read official documents and conducted interviews with officers from the Ministry of Education. I studied the reports of the two large scale projects recommended to me by Ministry officials and traced their origins and outcomes in policy. What follows is a summary of some of the findings and discussion of some of the issues.

8.1 Origins

A recent increase in interest in the process of educational evaluation has been driven by the emphasis on accountability within public and private management as well as in the education field as a whole. Evaluation of issues in education systems began in the form of consideration by groups of experts on public commissions. This system, which lasted for

many years, relied on informed opinion and concern for public acceptance rather than systematic investigation. Evaluation as a process has now developed its own methodologies with its own ethical standards and has become more tightly connected to the development of government policy in all social areas, including education. This has meant that rather than using the findings of existing research to inform policy, the Ministry of Education has taken greater control, both by initiating its own innovative programmes and for the evaluation of these.

Programme evaluation has developed its own schools of thought. Like research in general, there are differences of opinion over both the appropriate methods and the purpose of programme evaluation. All that seems to be agreed is that evaluation seeks a justification for existing practices and for interventions with the goal of finding ways of improving practices and procedures that will lead to improved outcomes. One thing however that comes through discussions is the concern of evaluators that their findings be used. Patton's utilisation-focused evaluation is an attempt to overcome this problem. Utilisation-focused evaluation is a process for helping the intended users and stakeholders to select the most appropriate content, model, methods, theory, and uses for their particular situation in the belief that this will ensure the use of the results.

Formal educational programme evaluation has been identified as beginning in America in 1887 with Joseph Rice's evaluation of spelling. Programme evaluations were undertaken from the 1970s by the New Zealand Department of Education (now replaced by the Ministry of Education). It is only more recently that these efforts have been identified as "programme evaluations" meaning a formal judgment of a project or practice within institutions using the techniques of research. Whether it was called programme evaluation or not, there does not appear to be a systematic history of such an activity within New Zealand, much less its deliberate use in informing policy. Those I interviewed about its use within the Ministry of Education were not able to talk about times outside the period of their tenure. The institutional memory of programme evaluation in education therefore appears very short. New Zealand's short experience of formal programme evaluation means that there has been, and still is, considerable reliance on international research to discover "what works" in education. But the comment has been made that:

> It is one thing to assert that the study of foreign education... [is] a valuable enterprise; it...[is] quite another to believe that foreign examples could be imported and domesticated. (Noah & Eckstein, 1969, p. 21)

In 1987 a Treasury briefing paper on education was presented to the incoming government. This paper stands as a landmark in educational policy development. The Treasury and the State Services Commission, the control departments within the public service, had developed a proposal for education in New Zealand based on both educational research and economic theory. In the briefing paper, the Treasury used existing research and theory to support its case for free market principles to be applied to all levels of education. This document came under sustained criticism from educational researchers who found their papers being used for purposes different from those for which they were intended. Although the briefing was a landmark in educational policy development, and led to Tomorrow's Schools, the policy changes recommended used available research. It did not generate new research, and was highly selective in what it chose to report. Its recommendations applied to the education system as a whole and not to what went on in classrooms, although there was the assumption that introducing free market principles into education would, in fact, raise educational standards. More significantly it was not prepared by the Department of Education as representing the education sector.

8.2 Outputs and Outcomes

The change to the specification of research and evaluation according to the needs identified by government departments and the letting of contracts for the prosecution and evaluation of such research came with the shift from outputs to outcomes as the management goal of the government.

Outputs are defined as the programmes that ministries, departments and other agencies are asked by Ministers to provide; whereas outcomes are the results that the policies actually deliver. The New Zealand Government appears to have adopted the British position that policy making is ultimately about delivering outcomes. It was explained to me that, from this point of view, governments demand solid evidence of effectiveness in programmes and practices in schools, on the understanding that action resulting from the evidence will see genuine progress, instead of the more familiar pendulum-like swings of opinion and

fashion. Ministers may want to focus on what is genuinely effective rather than what is fashionable, feels good, or possibly only works in theory. All divisions of the Ministry of Education are now involved with the linking of evaluation to policy. This has led to the planning of a large number of programme evaluations, determining the availability of expert knowledge, and finding persons able to carry out appropriate evaluations. The process is inevitably far from simple. The need to meet the needs of stakeholders who may not share the same goals, and the issue of the distribution of power within the evaluation process, have especial relevance for programmes affecting Māori pupils and for their evaluation.

8.3 The OECD Report

The 1989 New Zealand education reforms, which continue to the present, have led to significant changes in curriculum, governance, and management in schools. However, an OECD report in 2001 expressed the view that the changes in policy made prior to 2000 were driven by political conviction rather than evidence-based analysis with no systematic evaluation of their impact. Conviction politics sees no role for research or programme evaluation in policy determination. The OECD examiners also gained the impression that the great bulk of educational research in New Zealand was at the "applied end". They also believed that no active integration between research and policy existed at any level of the education system.

However, this is not the view of the Ministry of Education, which by 2000 had implemented collaborative interventions that did not look at the results of programme evaluation in isolation but considered their role in planning. Cycles of self-review had also been introduced and these had built a stronger relationship between programme evaluation and policy.

8.4 Policy Targets

Once it is accepted that formal investigations can provide information useful for policy development, programmes can be developed to find solutions to current problems. The demographic predictions are that in the year 2040 New Zealand will have a white minority population. If this prediction is correct, the present Māori and Pasifika students, the ones who make up the bulk of those who are being failed by the present education system, will

dominate the New Zealand workforce from 2040. The issue of raising the achievement of the current student population, especially those in “the tail” (a portion of the student body achieving well below international standards) needs addressing now. The presence of this "tail" of underachieving students in classrooms has resulted in the government's attention shifting to ways of resolving the problem. Evaluation of programmes should follow as swiftly as possible after the report of the results and there is a strong argument for formative evaluations that allow for correction as a trial programme proceeds. The current emphasis is on the role of teachers and their professional development as the means of raising achievement

It can be seen that the influences on policy are not restricted to local conditions but are influenced by international competition and results from international studies of achievement. A further finding which has been influential is that there is a greater difference in achievement levels within schools/classrooms than between schools. This was reported in both the 1970 International Association for Evaluation of Educational Achievement (IEA), and the Performance for International Student Assessment (PISA) (2000).

8.5 Two Ministry of Education Programme Evaluations

When I asked the Ministry of Education to recommend two programme evaluations for study, they recommended the SEMO project and The Numeracy Development Project both of which involve intervention and evaluation. The first is concerned with raising achievement in literacy and the second with raising achievement in numeracy. Both employ the strategy of professional development for teachers. These two projects have been described in chapters six and seven. In this present chapter they will be looked at as part of a sequence of policy development. One of the Government's strategic outcomes is to raise student achievement levels. Both projects address this problem. They focus on the raising student achievement. The outcomes reported have been favourable and evidence of these has been incorporated into policy as well as practice in an increasing number of schools.

The SEMO Project

On the basis of the SEMO results it was reported that teachers and principals lacked the knowledge of assessment, data analysis, and analysis use. These issues are neither new nor confined to the Mangere/Otara area but they led to implementation of the AUSAD (Analysis and Use of Student Assessment Data) initiative within the SEMO intervention. Analysis and use of assessment data has now become a national focus.

Further, the SEMO project used the professional development initiative called Picking up the Pace. The SEMO reports said that Māori and Pasifika students from low socio-economic backgrounds and in decile 1 schools can achieve the same outcomes as any other six-year-old. Raising teachers' generally low expectation for decile 1 students is claimed as a key to sustained student achievement.

Because this claim is based on research and evaluation rather than "political conviction" or ideology, it is open to challenge on the grounds of its research claims. The original Picking up the Pace study has been challenged. The Massey University Education Policy Group's (2003) analysis concluded that the numbers in the study were too small to justify the claims made and the policy conclusions drawn from it. One third of the children left the schools during the course of the study, and the control groups were such that no "cause and effect" relationships could be established. The researchers were criticised on the grounds that they assumed that because the students attended a decile 1 school they all had the same backgrounds. Most of the children studied were still in the bottom quartile on most of the tests after three years. The Education Policy Group (2003) critics say that this makes any claim that they were performing at "normal" levels for New Zealand six-year-olds unsupportable. In addition, the final group of teachers studied varied from those who began the original study. Therefore there was an assumption that the second teacher had made the same changes and progress as the first teacher would have, had he or she remained in the job.

The evaluation was criticised on the grounds that the data were presented in pseudo-quantitative ways when they were actually qualitative. Further, it appeared that the evaluators used only key findings to draw their conclusions, rather than looking at all their evidence. It is worth remembering Toulmin's (cited in Gorard, 2002) warning, "Producing

high quality research is important but even high quality work can lead to inappropriate conclusions" (p. 136). The OECD (2001) made a similar comment,

> New Zealand is not particularly lacking in the amounts of data… Arguably, there is too much data to be satisfactorily handled; for example the accumulation of data on children in relation to school accountability seems to have reached saturation levels (evidence from SEMO)…but there is little capacity, even amongst established academic researchers, for exploring datasets systematically. (p. 18)

In recent media reports on literacy, the Ministry cites the Picking up the Pace research as part of the justification for the expenditure of some ten million dollars. The evaluators explained their results on the theory that raising teacher expectations of their students' ability is the key to raising student achievement. Critics have asked whether this particular theory is well founded and have challenged the Minister of Education and Ministry's unquestioned acceptance of the findings. Based on current evidence, the initiative would benefit from an independent programme evaluation to justify the expenditure. A further issue, referred to in the literature on evaluation, is that one of the creators of the Picking up the Pace programme was also one of its evaluators which leads to considerations of partiality or bias. Debate continues in the *New Zealand Journal of Educational Studies.*

Picking up the Pace involves additional attention being given to literacy. There is a need for an evaluation, independent of those previously carried out, to ascertain the effect, if any, the Picking up the Pace programme has had on student achievement in curriculum areas other than literacy. Anecdotal evidence from teachers indicates that the demands the programme places on teachers means there is limited time to maintain a balanced curriculum. This would be a case of the solution to one problem creating another.

The Numeracy Development Project

The Numeracy Development Project is a nation wide project and focused on only one aspect of the curriculum, but the problems of low student achievement and a need to increase teacher knowledge were similar to those in the SEMO intervention.

Facilitators working with individual teachers in their classrooms were able to build a link with the intended programme, and help teachers to modify their pedagogical practice. The Ministry supports facilitation as the key variable in bridging the dissemination gap between researchers and practitioners and in the Numeracy Development Project facilitator quality was shown to influence the level of teacher up-take and student achievement.

The Numeracy Development Project programme evaluation was of a formative nature employing Rist's (1998) policy development cycle. Following this approach, qualitative investigation informed each of the three phases of *formulation*, *implementation* and *accountability* as policy intentions were translated into policy outputs and consequent outcomes. This illustrates policy as an iterative process involving the mathematics community. It is an approach that is distinctly better for Māori than approaches which do not encourage policy input from the beginning of the process because it enables Māori perspectives and aspirations to be clarified before it becomes too late to integrate them.

The Learning Policy Frameworks Team of the Ministry was responsible for the programme implementation. The evaluation involved both policy and curriculum personnel. This combination of curriculum interests with policy ensured that knowledge and pedagogy were a focus of the development. It is also an illustration of the collaboration needed to carry projects through. Free access to a report of the programme and its resources are on the Ministry of Education website, and this has increased the knowledge of teachers not yet involved in the intervention.

8.6 Outcomes in Policy

The Flaxmere Project, East Coast Project, Far North Project, West Coast Project and the Achievement Porirua Initiative have been, or are in, the process of being implemented as policy. These are similar to the SEMO intervention in that they were designed to reduce disparity and to raise student achievement in schools in demographic areas similar to Mangere and Otara. They build on the reports of the findings from the SEMO programme evaluations and use similar methods.

It would be wrong to suppose a simple journey between favourable results from the evaluation of an intervention and subsequent policy decisions. There may be barriers to

implementation that lie in the legislative controls on the education system. The earlier education reforms had led to legislation which prevented the Ministry of Education from playing a hands-on role in schools. Previously the Minister had been restricted in the action he/she could take to limit problems in the Mangere/Otara schools. This problem was addressed in the passing of the Education Bill 2000 that gave the Minister of Education more power to intervene when a school showed signs of being at risk. Also the National Education Guidelines were revised in 2000 to identify numeracy and literacy as priorities for all students, Years 0-4. Thus the regulatory structure was altered to allow intervention of a limited kind. As a consequence the Ministry of Education's perspective on school accountability has changed. The right of a school's operational autonomy is retained but an obligation to make a difference in teaching quality by reviewing practices was added to the 2004 Reporting and Planning requirements. These now demand evidence of how schools are trying to reduce the disparity between outcomes and expectations. It is no longer sufficient to provide only evidence of outcomes. This requirement sends clearer signals to schools of the importance of reporting collective achievement and emphasises review and strategic planning. The requirement is supported by an interactive website, "Leadspace" and there is a contract to train Board of Trustee members in compliance.

The Ministry responded to the SEMO finding of teachers' lack of assessment data analysis by developing the National Assessment Tools: the Exemplars in all curriculum areas, "asTTle" (Assessment Tool for Teaching and Learning) produced in English and Māori, in literacy and mathematics. These are to assist schools improve the quality of assessment of students, and to develop teacher capacity in curriculum knowledge and pedagogy. The ABeL (Assessment for Better Learning) and AtoL (Assess to Learn) professional development contracts for formative assessment run alongside the development of the tools. The Ministry of Education states that the gathering and analysis of high-quality data and evidence of student achievement, and the use of externally referenced exemplars are already proving to be powerful tools to influence policy development and implementation, which in turn lead to higher student achievement. The longer funding periods (2-3 years) for professional development contracts such as the AtoL Initiative and longer adviser contact with each school gives greater opportunity for the embedding of knowledge and skills. This is evidence of a change from familiar "hit and run," short-funded periods of professional development resulting in limited increase in knowledge or skill.

The plans are for the Numeracy Development Project to be extended to all students from Year 0-10.

8.7 Evaluators

Evaluators, Mansell, Renwick and Gray (1997), contracted to the Ministry of Education have commented that most funding agencies like good news and the political nature of programme evaluation contracts can result in rushed evaluations, watered-down results and recommendations, and inappropriate findings. They experienced severe limitations on the time available for them to carry out an evaluation when reviewing the implementation of compressed primary teacher education programmes. Contractors are likely to find themselves working in areas of dispute. In a paper with the title "Damned if we do and damned if we don't: dilemmas of research on contentious policy", they outline the problems involved in establishing adequate evaluation procedures for a new government policy. There is often a lack of the baseline data that is needed for comparisons or monitoring. Assumptions of causality and effect are problematic; for example, how to isolate and determine the contribution one programme has in the life of a student.

Evaluators see some schools, especially decile 1, as having so many programmes being implemented and evaluated that they have someone regularly observing in their classes. Interpretations of results tend to overlook the Hawthorne Effect, however, where productivity is increased in a worker with the psychological stimulus of being singled out and made to feel important.

Evaluators as informants referred to positive moves being made by the Research Division within the Ministry, especially in driving the use of research in the development of policy, but that there are still many issues to work through before a comfortable and equitable relationship is established between those responsible for research and those responsible for policy. This point was emphasised also in my interviews with Ministry officers.

Whereas previously, evaluators were likely to be researchers rather than evaluation specialists it is clear from the record of those who have had evaluation contracts from the Ministry of Education that a core of evaluators, drawn largely from the education departments of universities, is now emerging. The skills required by an evaluator are not

identical with those needed by a pure researcher. Evaluators, as the account of the SEMO project in Chapter Six showed, need moderation skills to bring various stakeholders together and to gain commitment to a common purpose. Within the Numeracy Development Project, the evaluators addressed the needs of all stakeholders: teachers, students, facilitators, programme developers and policy-makers.

There continue to be issues related to the skills of an evaluator and controversy over whether an evaluator needs to be a professional in the field relevant to the programme being evaluated. Problems may arise, equally, if a sole evaluator is also a specialist in the field being evaluated. On the other hand it could be argued that bias may result if the evaluator is not familiar with the field. There may be no final answer, the situation depending ultimately on the nature of the programme and the evaluator's experience. Given the small size of the education community in New Zealand, a further question is whether an evaluator should be one of the authors of the development of the programme under evaluation and, on the other hand, how formative evaluation can be carried out without involving the authors of an intervention. In view of questions such as these it may be time for evaluators of social programmes in New Zealand to adopt a code of ethics along the lines of the code developed by their American counterparts. Also there is a shortage of courses of training in programme evaluation including training in evaluation carried out under contract.

The SEMO report describes an attempt to develop appropriate methods of evaluation and the Ministry of Education has selected and promoted methods such as intervention logic as a means of arriving at sound results.

8.8 Programme Evaluation and the Performance Based Research Fund (PBRF)

The first PBRF ranking of research has taken place and the results released. In New Zealand, institutions such as colleges of education have substantial contracts for programme evaluation but not for "pure" research. Interviews with Ministry of Education personnel showed that, within the Ministry, programme evaluation is equal in status to pure research. However, the universities, whose members dominate the panels judging the quality of research, do not appear to rank research and evaluation as equal. Given that universities undertake over half New Zealand's educational research, there is a potential issue here. At a time when the

government is trying to increase the capability and capacity of educational researchers there appears to be limited academic recognition or funding for the type of investigations that are likely to have application in the determination of policy. This could lead to university staff being discouraged from undertaking Ministry of Education contracts. To deepen the problem, most of the colleges of education which are contracted to undertake programme evaluations for the Ministry are now either merged, or in the process of merging, with universities. This has implications for the future of the research status of universities and particularly their education departments.

It can be seen that the government's decision to produce outcomes, and the deliberate encouragement of applied research and evaluation by the Ministry responsible for educational policy have had downstream effects on institutions whose financial support and prestige depend in part on rather different styles of research.

8.9 Conclusion

This thesis has covered a number of different topics in an exploration of programme evaluation and its place in the processes of policy determination. The technical methods of programme evaluation are still developing, its place within the system of contracting operated by the Ministry of Education will continue, and while overseas research is still useful in determining what works in education, local projects have the advantage of letting us see whether they work in practice. Programme evaluation has a firm place in continuing efforts to improve education. While this thesis has done no more than open up the field in which programme evaluation is situated it has demonstrated both some of its uses, and it usefulness in the implementation of government policy. The topic deserves further investigation.

REFERENCES

Alton-Lee, A. (2004). *Guidelines for generating a best evidence synthesis iteration.* Wellington: Ministry of Education.

Annemalm, C., & Nilsson, K. (2003). *Programme logic.* Sweden: ecs ekonomistrningsverket.

Austin, M. (1988). Issues and practicalities: The government's response to Picot. In C. Wylie (Ed.), *Proceedings of the First Research into Educational Policy Conference* (pp.1-10). Wellington: NZCER.

Bailey, J. P. M. (1995). *An evaluation of compulsory breath testing in New Zealand.* Porirua: Institute of Environmental Science and Research.

Baker, R. (2000). Untitled paper presented at the Regional Seminar, *Educational research for policy and practice: With particular reference to secondary education reform,* Tokyo 2-7 October. Retrieved October 16, 2003, from http://www.nzcer.org.nz

Ball, S. (1990). *Politics and policy making in education.* London: Routledge.

Bauman, A., McLean, G., Hurdle, D., Walker, S., Boyd, J., van Aalst, I., & Carr, H. (2003). Evaluation of the national "Push Play" campaign in New Zealand: Creating population awareness of physical activity. *New Zealand Medical Journal,* 116, 1179.

Beal, F. (2003). *Balancing demands: Exploring dilemmas that can arise in evaluative research.* Paper presented at the AARE/NZARE Conference, Auckland, November 30 – December 3.

Boulding, K. E. (1980). Science: Our common heritage. *Science,* 207, 831-836.

Brandl, J. (1994, September 5). Must deal with the bureaucracy – But exactly how is harder to say. *Minneapolis Star Tribune,* A13.

Cabinet Office. (1999). Modernising government. In S. Nutely, I. Walter & H. Davies (2002). *Evidenced based policy and practice: Cross sector lessons from the UK.* St. Katharine's West: University of St. Andrews.

Campbell, D.T. (1972). *Methods for the experimenting society.* A paper presented at the annual meeting of the American Psychological Association, Washington, D.C.

Chateris, P. L. (2002). *Genetic change.* Palmerston North: Institute of Veterinary and Biomedical Services, Massey University.

Chelimsky, E. (1995). The political environment of evaluation and what it means for the development of the field. *Evaluation Practice, 16* (3), 215-225.

Chelimsky, E. (1997). The coming transformations in evaluation. In E. Chelimsky & W. Standish (Eds.), *Evaluation for the 21st century.* Thousand Oaks, CA: Sage.

Cronbach, L. J. (1982). *Designing evaluations of educational and social programs.* San Francisco: Jossey-Bass.

Coe, R. (2002). *Finding out what works: Evidence-based education.* Retrieved October 17, 2002, from http://www.cem.dur.ac.uk/ebeuk

Denscombe, M. (1998). *The good research guide for small-scale social research projects.* Buckingham: Open University Press.

Denzin, N.K., & Lincoln, Y.S. (Eds.). (2000). *Handbook of qualitative research* (2nd ed.). Thousand Oaks, CA: Sage.

Department of Education. (1978). *The first 100 years.* Wellington.

Editorial. (2003, April 24). Are we getting value for money? *The Dominion Post,* B4.

Education Policy Group. (2003). Literacy research in South Auckland: A critique. *Input, 25* (3), 14-15.

Education Review Office. (1994). *Mathematics in the New Zealand curriculum.* Wellington.

Education Review Office. (1996). *Improving schooling in Mangere and Otara.* Wellington.

Education Review Office. (1998). *Schooling in Mangere and Otara: Progress since 1996.* Wellington.

Fetterman, D. (1997). *A response to Patton and Scriven's review of empowerment evaluation*. Retrieved October 15, 2002, from http://www.stanford.edu/~davidf/pattonscriven.html

Flockton, L., & Crooks, T. (1997). *Mathematic assessment results, 1997.* National Education Monitoring Report 9. Dunedin: Educational Assessment Research Unit.

Floden, R., & Weiner, S.S. (1978). Rationality to ritual: The multiple roles of evaluation in government processes. *Policy Sciences*, *9* (9), 9-18.

Funnell, S. (1997). Program logic: An adaptable tool for designing and evaluating programs. *Evaluation News and Comment*, July.

Gorard, S. (2002). Fostering skepticism: The importance of warranting claims. *Evaluation and Research in Education 16* (3), 136-149.

Gordon, L. (1993). From professional to proletarian: Teachers and the state Aotearoa. *Delta,* 47, 33-46.

Greene, J. C. (1994). Qualitative program evaluation: Practice and promise. In N. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research.* Thousand Oaks, CA: Sage.

Grimes, A. (2003). *Annual review*. (SPEaR Social Policy Evaluation and Research Committee). Wellington: Ministry of Social Development.

Guba, E.G., & Lincoln, Y.S. (1989). *Fourth generation evaluation.* Newbury Park, CA: Sage.

Hammersley, M. (2002). *Educational research: Policy making and research.* London: Paul Chapman Publishing.

Higgins, J. (2001). *An evaluation of the year 4-6 numeracy exploratory study: Exploring issues in mathematics education.* Wellington: Learning Media.

Higgins, J. (2002). *An evaluation of the advanced numeracy project 2001: Exploring issues in mathematics education.* Wellington: Learning Media.

Higgins, J., Parsons, R., & Hyland, M. (2003). The numeracy development project: Policy to practice. In *The New Zealand Annual Review of Education 2003.*

Higgins. J. (2003). *An evaluation of the advanced numeracy project 2002: Exploring issues in mathematics education.* Wellington: Learning Media.

Higgins, J. (2003). (pers. com). *Te Kauhua: Māori in the mainstream pilot project.* A discussion of evaluation findings with staff, Wellington College of Education.

Holton, D., Spicer, T., Thomas, G., & Young, S. (1996). *The benefits of problem solving in the learning of mathematics.* Final Report. Wellington: Ministry of Education, Research and International Section.

House, E. (1993). *Professional evaluation: Social impact and political consequences.* Newbury Park, CA: Sage.

Hume, W., & Bryce, T. (2001). Scholarship: Research and the evidential basis of policy development in education. *British Journal of Educational Studies, 49* (3), 329-352.

Irving, J. (1992). *Research/policy interface.* An unpublished paper presented at the Policy Refocusing Working Group, Wellington.

Irwin, K. (2003). *An evaluation of the numeracy exploratory study years 7-10, 2002: Exploring issues in mathematical education.* Wellington: Learning Media.

Irwin, K., & Niederer, K. (2002). *An evaluation of the numeracy exploratory study years 7-10, 2001: Exploring issues in mathematical education.* Wellington: Learning Media.

King, J. A., & Pechman, E. M. (1984). Pinning a wave to shore: Conceptualising school evaluation use. *Educational Evaluation and Policy Analysis, 6* (3), 241-51.

Lau, G. (1997). Changing roles: Evaluator and teacher collaborating in school change. *Evaluation and Programme Planning, 20* (1), 7-15.

Little, W. Fowler, H. W., & Coulson, J. (1973). (prepared by). *Shorter Oxford English dictionary on historical principles* (3rd ed.). Oxford: University Press.

Livingstone, I. (1989). Director's comment in the *54th Annual Report, NZCER.* Wellington: New Zealand Council for Educational Research.

Lomas, J. (2000). Connecting research and policy. *isma,* 1(1).

Madaus, G., Scriven, M., & Stufflebeam, D. (Eds.). (1983). *Evaluation models: Viewpoints on educational and human services evaluation.* Boston: Kluwer-Nijhoff Publishing.

Mansell, R., Renwick, M., & Gray, A. (1997). *Damned if we do and damned if we don't: Dilemmas of research on contentious policy.* A paper presented to the New Zealand Association for Research in Education Conference, University of Auckland, 6 December 1997.

Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding and improving public and nonprofit policies and programs.* San Francisco: Jossey-Bass.

McDonald, G. (1993). Research and government policy in New Zealand. *Evaluation and Research in Education, 7* (1), 11-21.

McLaughlin, J.A., & Jordon, G. B. (1999). Logic models: A tool for telling your programmes performance story. *Evaluation and Program Planning*, 22, 65-72.

McNamara, C. (1998). *Basic guide to programme evaluation.* Retrieved August 2 2002, from http://www.mapnp.org/library/evaluatn/fnl_eval.htm

McMahon, T. (2002). *Evidence in action.* Paper presented at National Assessment Regional Seminars, April. Retrieved June 3, 2003, from http://www.tki.org.nz/r/assessment/research/research15_e.php

McPherson, A., & Raab, C. D. (1988). *Governing education: Sociology of policy since 1945.* Edinburgh: Edinburgh University Press.

Middleton, S., Codd, J., & Jones, A. (Eds.). (1990). *New Zealand education policy today: Critical perspectives.* Wellington: Allen & Unwin.

Ministry of Education. (1988). *Tomorrow's schools: The reform of education administration in New Zealand.* Wellington.

Ministry of Education. (1992). *New Zealand national mathematics curriculum.* Wellington.

Ministry of Education. (1997). *The mathematics and science taskforce: The Report.* Wellington.

Ministry of Education. (1999). *Report of the literacy taskforce.* Wellington.

Ministry of Education. (2000). *The national education guidelines.* Wellington.

Ministry of Education. (2001). *Mapping the building of capacity and capability within the educational research community.* Retrieved October 21, 2002, from *http://www.minedu.govt.nz*

Ministry of Education. (2002). *2001-2002 annual report*. Wellington.

Ministry of Education. (2003a). *Statement of intent: 2003 - 2004.* Retrieved May 17, 2003, from http://www.minedu.govt.nz/index.cfm?layout

Ministry of Education. (2003b). *The number framework: Draft teachers' materials.* Wellington: Learning Media.

Ministry of Social Development. (2003). *Annual review: SPEaR social policy evaluation and research committee.* Wellington.

Mitchell, L., & Mara, D. (2001). *Evaluation of two ECD services: Licensing and chartering advice and support: Advice and support to license-exempt playgroups.* Report to the Ministry of Education. Wellington: Ministry of Education.

Morris, L.L., & Fitz-Gibbon, C.T. (1978). *How to deal with goals and objectives.* Beverley Hills, CA: Sage.

Nash, R. (2004). *One PACE forward two steps backwards: Reading achievement, teacher expectations and educational policy.* A Draft paper.

New Zealand Government. (2002a). Guidance on outcomes based management. In *Managing for outcomes: Guidance for departments:* A report of the steering group, Managing for Outcomes Roll-out, 2003/04. Wellington.

New Zealand Government. (2002b). *Delivering improved results to New Zealanders. Pathfinder outcomes guidance – version 1.2.* Wellington: Ministry of Education.

Noah, H.J., & Eckstein, M.A. (1969). In D. Phillips & K. Ochs. (2004). Researching policy borrowing: some methodological challenges in comparative education. *British Educational Research Journal, 30* (6). 772-784.

Nutely, S., Walter, I., & Davies, H. (2002a). *From knowing to doing: A framework for understanding the evidence – into – practice agenda. Discussion Paper 1.* St Katharine's West: Research Unit for Research Utilisation, University of St. Andrews.

Nutely, S., Walter, I., & Davies, H. (2002b). *Evidenced based policy and practice: Cross sector lessons from the UK.* St Katharine's West: University of St. Andrews.

O'Neill, H. (2002). *Called to account*. BBC Reith lectures 2002.

OECD Review. (2001). *Educational research and development policy in New Zealand.* Examiners' Report. CERI/CD(2001)4.

Patton, M.Q. (1990). *Qualitative research and methods* (2nd ed.). Newbury Park, California: Sage Productions.

Patton, M.Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.) Thousand Oaks, California: Sage Publications.

Patton, M.Q. (2001). Evaluation, knowledge management, best practices, and high quality lessons learned. *American Journal of Evaluation, 22* (3), 329-336.

Patton, M.Q. (2002). *The Evaluation Exchange Newsletter*. *Family Support Issue.* Spring 2002. Retrieved October 15, 2002, from http://www.gse.harvard.edu/~hfrp/eval/issue18/qanda.html

Pawson, R. (2001a). *Evidenced based policy:11. The promise of realist synthesis* (Working Paper 4). London: ESRC, Queen Mary University.

Pawson, R. (2001b). *Evidenced based policy and naming and shaming.* (Working Paper 5) London: ESRC, Queen Mary University.

Phillips, G. E., McNaughton, S., & MacDonald, S. (2002). *Picking up the pace: Effective literacy interventions for accelerated progress over the transition into decile one schools* (Final Report). Wellington: Ministry of Education.

Ramsay, G. (1981). *It may be too late* (Report to the Department of Education). Wellington: Department of Education.

Rist, R. C. (1998). Influencing the policy process with qualitative research. In N.K. Denzin, & Y.S. Lincoln (Eds.) (2000), *Handbook of qualitative research* (2nd ed.). Thousand Oaks, CA: Sage Publications Inc.

Robinson, V., Timperley, H., & Bullard, T. (2000). *Strengthening education in Mangere and Otara evaluation: Second evaluation report.* Wellington: Ministry of Education.

Robinson, V., & Timperley, H. with Ward, L., Tu'uga Stevenson, V. & Mitchell, S. (2004). *Strengthening education in Mangere and Otara evaluation.* Final Evaluation Report. Wellington: Ministry of Education.

Robson, C. (1993). *Real world research: A resource for social scientists and practitioner-researchers.* Oxford: Blackwell.

Rogers, P. (2003). *Pathways to the future: Nga Huarahi Arataki.* Melbourne: Royal Melbourne Institute of Technology.

Scriven, M. (1983). Evaluation ideologies. In G. Madaus, M. Scriven & D. Stufflebeam (Eds.), *Evaluation models viewpoints on educational and human services evaluation* (pp. 229-260). Boston: Kluwer-Nijhoff Publishing.

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.

Scriven, M. (1993). Hard-won lessons in program evaluation. *New Directions for Program Evaluation.* 58.

Shragge, E., Church, K., Fontan, J., & Lachance, E. (1999). *Alternative/training businesses: New practice, new directions.* Canada: Human Resources Development.

Slavin, R.E. (1987). Ability grouping and student achievement in elementary schools: A best evidence synthesis. *Review of Educational Research, 60* (3), 471-499.

Slavin, R.E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher, 31* (7), 15-22.

Smith, N. (1980). Studying evaluation assumptions. *Evaluation Network Newsletter,* Winter, 39-40.

Smith, P. (2002). *An academic's view on influencing health policy.* A paper presented at How can health economists outside of DN impact on health policy? OHE seminar, London, 13 September.

Stake, R. (1967). Toward a technology for the evaluation of educational programme. *AERA Monograph Series: A Curriculum Evaluation,* 1, 1-12.

Stake, R. (1995). *The art of case research.* Newbury Park, CA: Sage.

Stake, R. (1998). When policy is merely *promotion, by what ethic lives an evaluator? Studies in Educational Evaluation, 24* (2), 203-212.

Stone, D. (1997). *The policy paradox: The art of political decision-making.* New York: W.W. Norton.

Stoney, S. (2000). *Dragons, demons and angels: Programme evaluation at NFER down the years.* A paper presented 4 October 2000 to an unknown audience.

Stufflebeam, D.L. (1994). Empowerment evaluation, objectivist evaluation, and evaluation standards: Where the future of evaluation should not go and where it needs to go. *Evaluation Practice, 15* (3), 321-38.

The Treasury. (1987). *Government Management: Brief to the Incoming Government 1987 Vol.11, Education issues.* Wellington: The Treasury.

Thomas, G., & Ward, J. (2001). *An evaluation of the count me in too pilot project: Exploring issues in mathematics education.* Wellington: Learning Media.

Thomas, G., & Ward, J. (2002). *An evaluation of the early numeracy project 2001: Exploring issues in mathematics education.* Wellington: Learning Media.

Thomas, G., Tagg, A., & Ward, J. (2003). *An evaluation of the early numeracy project 2002: Exploring issues in mathematics education.* Wellington: Learning Media.

Timperley, H. S., & Lam, P. (2002). *Analysis and use of student achievement data. Baseline report.* Wellington: Ministry of Education.

Timperley, H., Phillips, G., & Wiseman, J. (2003). *The sustainability of professional development in literacy, Part 1: Changing and sustaining teachers' expectations through professional development in literacy, Report to the Ministry of Education.* Wellington: Ministry of Education, Research Division.

Timperley, H., Robinson, V., & Bullard, T. (1999). *Strengthening education in Mangere and Otara: First evaluation report.* Wellington: Leaning Media.

Tuuta, M., Bradnam, L., Hynds, A., & Higgins, J. with Broughton, R. (2002). *Te Kauhua: Māori in the mainstream pilot project.* Wellington: Wellington College of Education.

Tyler, R. (1983). In G. Madaus, M. Scriven & D. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation.* Boston: Kluwer-Nijhoff Publishing.

University of Otago, (2003). *Performance-based research fund (PBRF).* Retrieved October 27, 2004, from http://www.otago.ac.nz/research/pbrf/

Visser, H. (Ed.). (1999). *Exploring issues in mathematics education.* Wellington: Learning Media.

Waters, G. (1998). Critical evaluation for education reform. *Education Policy Analysis Archives, 6* (20), 1-37.

Weiss, C. (Ed.). (1972). *Evaluating action programs.* Boston: Allyn & Bacon.

Whyte, G. (1989). In M. Q. Patton (1990), *Qualitative research and methods* (2nd ed.). Newbury Park, California: S age Productions.

Wolfe, T.H. (1973). Alfred Binet. Chicago: University of Chicago Press.

Young-Loveridge, J., & Wright, V. (2002). Validation of the N.Z. number framework. In B. Barton, K. Irwin, M. Pfannkuch, & M. Thomas (Eds.), *Mathematics education in the South Pacific* (Vol. 2). Proceedings of MERGA 25, Auckland.

Appendices

APPENDIX ONE

Informed Consent Form

I _____ give my consent for Helen Jackman, Wellington College of Education, to use information collected in an interview with me at the Ministry of Education on Thursday 13th March, 2003.

I understand the information will be published in Helen's Masters thesis that will discuss educational programme evaluation and the interface between the Research and Policy sections of the New Zealand Ministry of Education.

All information pertaining to me will be presented for my verification before the final draft is compiled.

Confidentiality is assured and my consent may be withdrawn at any time up to the final publishing of the thesis.

Signed _____

Date _____

WELLINGTON
COLLEGE
OF EDUCATION
TE WHĀNAU
O AKO PAI KI
TE UPOKO O TE IKA

Box 17•310
Donald Street
Karori
Wellington
New Zealand

Telephone
0•4•924 2000

Facsimile
0•4•924 2079

Making a difference

APPENDIX TWO

January 21st 2003

I am enrolled in the Master of Education at Wellington College of Education. Currently I am preparing my thesis entitled, The Evaluation of Educational Programmes within New Zealand: a meta-evaluation.

I am a Senior Adviser in the School Support Services, Wellington College of Education and as such, have a deep interest in raising student achievement. The Ministry of Education has implemented many programmes and initiatives over the years to enhance student learning and has had them evaluated using varied approaches. I am especially interested in the use made of the findings of these evaluations.

In my initial use of programme evaluations it became clear that knowledge of the history of development and the processes used in the New Zealand Ministry of Education is necessary as a basis to understanding the current evaluation work. It would appear there is no documented history available at this point and I wish to begin filling that gap.

I believe that your experience and knowledge will be of particular assistance to me in preparing this thesis and would be grateful of any assistance you are able to offer.

The New Zealand Association for Research Code of Ethics will be adhered to at all times.

Thank you for your time and thoughts.

Helen Jackman
Senior Adviser
Wellington College of Education

WELLINGTON
COLLEGE
OF EDUCATION
TE WHĀNAU
O AKO PAI KI
TE UPOKO O TE IKA

Box 17•310
Donald Street
Karori
Wellington
New Zealand

Telephone
0•4•924 2000

Facsimile
0•4•924 2079

APPENDIX THREE
Interview Topics for discussion with Research and Evaluation Division

Questions	Prompt/Probe
When I analysed the lists I found that the earliest Ministry of Education's Research Report documented was a 1974 evaluation of the Pilot Schemes in Extended Courses for the First Year Apprentices carried out by the Department of Education. Many of the early reports listed pertain to areas such as carpentry and nursing. Who/What drove the development of programme evaluation in the MoE?	If they are unsure: is there anyone they could recommend who would have this information?
In preparing a template of the reports done pre and post 1989 it is evident there have been a great many more evaluations done post 1989. What are the significant changes in approach to programme evaluation have taken place post 1989? I understand you have attended the Evaluators' Institute Conference. What ideas were you able to implement from that into your MoE work?	If more formative: how do you manage this ongoing process within the time limits of government?
How does the Research Division manage an ongoing formative approach to programme evaluation?	
What processes are in place to strengthen the line between research and policy?	
How is programme evaluation funded and what part do policy play in the funding??	
How are the Research Division and policy teams meeting short, and long-term research needs that are based an both current policy needs and the provision of specific background information?	
What processes are in place to ensure there is dialogue between the Ministry of Education and other ministries that influence education policy?	
Does the Ministry of Education believe stakeholders should be involved in programme evaluation and policy development, and how could you ensure this happens given that you contract an evaluator in most circumstances?	
How does the MoE involve Māori stakeholders in the evaluations process?	
What part does programme evaluation play in policy development?	
What type of policy development does the Ministry of Education favour?	

Interview Topics for discussion with Learning Policy Frameworks Team

Questions	Prompt/Probe
Do you have short, medium and long-term research needs that are based on both current policy needs and the provision of specific background information?	
How do the IEA, PISA, and UNESCO etc results influence policy?	
What type of policy development does the Ministry of Education favour?	If evidence-based – What does it mean? What counts as evidence? Whose evidence? Why was it introduced? What kind of knowledge are you looking for?
James Irving believes the Ministry of Education policy analysts have had insufficient professional development in research, methodologies, and data analysis to ensure familiarity. What is your opinion on this statement?	
Do the Policy sections have anything to do with the allocation of research funding?	
What part does programme evaluation play in policy development?	
What processes are in place to ensure there is dialogue between the Ministry of Education and the Policy teams in other ministries?	What is the composition of Policy ?
To what extent are policy driven by the ultimate goals of the present government?	
The new Performance Based Research Funding Policy is going to have considerable implications for the tertiary sector. Are you optimistic about its influence on policy development?	What is going to count as research? If evaluation doesn't who will do their work?

The Number Framework - Strategies

	Global Stage	Operational Domains: Addition and Subtraction	Operational Domains: Multiplication and Division	Operational Domains: Proportions and Ratios
Counting	Zero: Emergent	**Emergent** The student is unable to count a given set or form a set of up to ten objects.		
	One: One-to-one Counting	**One-to-one Counting** The student is able to count a set of objects but is unable to form sets of objects to solve simple addition and subtraction problems.	**One-to-one Counting** The student is able to count a set of objects but is unable to form sets of objects to solve simple multiplication and division problems.	**Unequal Sharing** The student is unable to divide a region or set into two or four equal parts.
	Two: Counting from One on Materials	**Counting - from One** The student solves simple addition and subtraction problems by counting all the objects, e.g., 5 + 4 as 1, 2, 3, 4, 5, 6, 7, 8, 9. The student needs supporting **materials**, like fingers.	**Counting - from One** The student solves multiplication and division problems by counting one to one with the aid of **materials**.	**Equal Sharing** The student is able to divide a region or set into given equal parts using **materials**. With sets this is done by equal sharing of materials. With shapes symmetry (halving) is used.
	Three: Counting from One by Imaging	**Counting - from One** The student **images** all of the objects and counts them. The student does not see ten as a unit of any kind and solves multi-digit addition and subtraction problems by counting all the objects.	**Counting - from One** The student **images** the objects to solve simple multiplication and division problems, by counting all the objects, e.g., 4 × 2 as 1, 2, 3, 4, 5, 6, 7, 8. For problems involving larger numbers the student will still rely on materials.	**Equal Sharing** The student is able to share a region or set into given equal parts by using **materials** or by **imaging** the materials for simple problems, e.g., $\frac{1}{2}$ of 8. With sets this is done by equal sharing of materials or by imaging. With shapes symmetry is used to create halves, quarters, eighths, etc.
	Four: Advanced Counting	**Counting On** The student uses counting on or counting back to solve simple addition or subtraction tasks, e.g., 8 + 5 by 8, 9, 10, 11, 12, 13 or 52 – 4 as 52, 51, 50, 49, 48. Initially, the student needs supporting **materials** but later **images** the objects and counts them. The student sees 10 as a completed count composed of 10 ones. The student solves addition and subtraction tasks by incrementing in ones (38, 39, 40, ...), tens counts (13, 23, 33, ...), and/or a combination of tens and ones counts (27, 37, 47, 48, 49, 50, 51).	**Skip-counting** On multiplication tasks, the student uses skip-counting (often in conjunction with one-to-one counting), e.g., 4 × 5 as 5, 10, 15, 20. The student may track the counts using materials (eg. fingers) or by imaging.	

The Number Framework (Ministry of Education, 2003) pp. 9-16.

	Global Stage	Operational Domains: Addition and Subtraction	Operational Domains: Multiplication and Division	Operational Domains: Proportions and Ratios
Part-Whole	Five: Early Additive Part-Whole	**Early Addition and Subtraction** The student uses a limited range of mental strategies to estimate answers and solve addition or subtraction problems. These strategies involve deriving the answer from known basic facts, e.g., 8 + 7 is 8 + 8 – 1 (doubles) or 5 + 3 + 5 + 2 (fives) or 10 + 5 (making tens). Their strategies with multi-digit numbers involve using tens and hundreds as abstract units that can be partitioned, e.g., 43 + 25 = (40 + 20) + (3 + 5) = 60 + 8 = 68 (standard partitioning) or 39 + 26 = 40 + 25 = 65 (rounding and compensation) or 84 – 8 as 84 – 4 – 4 = 76 (back through ten).	**Multiplication by Repeated Addition** On multiplication tasks, the student uses a combination of known multiplication facts and repeated addition, e.g., 4 × 6 as (6 + 6) + (6 + 6) = 12 + 12 = 24. The student uses known multiplication and repeated addition facts to anticipate the result of division, e.g., 20 ÷ 4 = 5 because 5 + 5 = 10 and 10 + 10 = 20.	**Fraction of a Number by Addition** The student finds a fraction of a number and solves division problems with remainders mentally using halving, or deriving from known addition facts, e.g., $\frac{1}{3}$ of 12 is 4 because 3 + 3 + 3 = 9, so 4 + 4 + 4 = 12; e.g., 7 pies shared among 4 people (7 ÷ 4) by giving each person 1 pie, and $\frac{1}{2}$ pie, then $\frac{1}{4}$ pie.
	Six: Advanced Additive (Early Multiplicative) Part-Whole	**Advanced Addition and Subtraction of Whole Numbers** The student can estimate answers and solve addition and subtraction tasks involving whole numbers mentally by choosing appropriately from a broad range of advanced mental strategies, e.g., 63 – 39 = 63 – 40 + 1 = 24 (rounding and compensating) or 39 + 20 + 4 = 63, so 63 – 39 = 24 (reversibility) or 64 – 40 = 24 (equal additions) e.g., 324 – 86 = 300 – 62 = 238 (standard place value partitioning) or 324 – 100 + 14 = 238 (rounding and compensating).	**Derived Multiplication** The student uses a combination of known facts and mental strategies to derive answers to multiplication and division problems, e.g., 4 × 8 = 2 × 16 = 32 (doubling and halving), e.g., 9 × 6 is (10 × 6) – 6 = 54 (rounding and compensating), e.g., 63 ÷ 7 = 9 because 9 × 7 = 63 (reversibility).	**Fraction of a Number by Addition and Multiplication** The student uses repeated halving or known multiplication and division facts to solve problems that involve finding fractions of a set or region, renaming improper fractions, and division with remainders, e.g., $\frac{1}{3}$ of 36, 3 × 10 = 30, 36 – 30 = 6, 6 ÷ 3 = 2, 10 + 2 = 12 e.g., $\frac{16}{3} = 5\frac{1}{3}$ (using 5 × 3 = 15) e.g., 8 pies shared among 3 people (8 ÷ 3) by giving each person 2 pies and dividing the remaining 2 pies into thirds (answer: $2 + \frac{1}{3} + \frac{1}{3} = 2\frac{2}{3}$). The student uses repeated replication to solve simple problems involving ratios and rates, e.g. 2:3 → 4:6 → 8:12 etc.

Part-Whole

Global Stage	Operational Domains: Addition and Subtraction	Operational Domains: Multiplication and Division	Operational Domains: Proportions and Ratios
Seven: Advanced Multiplicative (Early Proportional) Part-Whole	**Addition and Subtraction of Decimals and Integers** The student can choose appropriately from a broad range of mental strategies to estimate answers and solve addition and subtraction problems involving decimals, integers, and related fractions. The student can also use multiplication and division to solve addition and subtraction problems with whole numbers. e.g., 3.2 + 1.95 = 3.2 + 2 – 0.05 = 5.2 – 0.05 = 5.15 (compensation); e.g., 6.03 – 5.8 = □, as 6.03 – 5 – 0.8 = 1.03 – 0.8 = 0.23 (standard place value partitioning) or as 5.8 + □ = 6.03 (reversibility) e.g., □ + 3.98 = 7.04 as 3.98 + □ = 7.04, □ = 0.02 + 3.04 = 3.06 (commutativity) e.g., $\frac{3}{4} + \frac{5}{8} = (\frac{3}{4} + \frac{2}{8}) + \frac{3}{8} = 1\frac{3}{8}$ (partitioning fractions) e.g., 81 – 36 = (9 × 9) – (4 × 9) = 5 × 9 (using factors) e.g., 28 + 33 + 27 + 30 + 32 = 5 × 30 (averaging) e.g., +7 – -3 = +7 + +3 = +10 (equivalent operations on integers)	**Advanced Multiplication and Division** The student chooses appropriately from a broad range of mental strategies to estimate answers and solve multiplication and division problems. These strategies involve partitioning one or more of the factors in multiplication, and applying reversibility to solve division problems, particularly those involving missing factors and remainders. The partitioning may be additive or multiplicative. e.g., 24 × 6 = (20 × 6) + (4 × 6) (place value partioning) or 25 × 6 – 6 (rounding and compensating) e.g., 81 ÷ 9 = 9, so 81 ÷ 3 = 3 × 9 (proportional adjustment) e.g., 4 × 25 = 100, so 92 ÷ 4 = 25 – 2 = 23 (reversibility and rounding with compensation) e.g., 90 ÷ 5 = 18 so 87 ÷ 5 = 17 r 2 (rounding and divisibility) e.g., 201 ÷ 3 as 100 ÷ 3 = 33 r 1, 200 ÷ 3 = 66 r 2, 201 ÷ 3 = 67 (divisibility rules)	**Fractions, Ratios, and Proportions by Multiplication** The student uses a range of multiplication and division strategies to estimate answers, and solve problems with fractions, proportions, and ratios. These strategies involve linking division to fractional answers, e.g., 11 ÷ 3 = $\frac{11}{3} = 3\frac{2}{3}$ e.g., 13 ÷ 5 = (10 ÷ 5) + (3 ÷ 5) = $2\frac{3}{5}$ The student can also find simple equivalent fractions, and rename common fractions as decimals and percentages. e.g., $\frac{5}{6}$ of 24 as $\frac{1}{6}$ of 24 = 4, 5 × 4 = 20 or 24 – 4 = 20 e.g., 3:5 as □ : 40, 8 × 5 = 40, 8 × 3 = 24 so □ = 24. e.g., $\frac{3}{4} = \frac{75}{100} = 75\% = 0.75$
Eight: Advanced Proportional Part-Whole	**Addition and Subtraction of Fractions** The student uses a range of mental partitioning strategies to estimate answers and solve problems that involve adding and subtracting fractions, including decimals. The student is able to combine ratios and proportions with different amounts. The strategies include using partitions of fractions and "ones", and finding equivalent fractions. e.g., $2\frac{3}{4} - 1\frac{2}{3} = 1 + (\frac{3}{4} - \frac{2}{3}) = 1 + (\frac{9}{12} - \frac{8}{12}) = 1\frac{1}{12}$ (equivalent fractions) 20 counters in ratio of 2:3 combined with 60 counters in ratio 8:7 gives a combined ratio of 1:1.	**Multiplication and Division of Decimals/ Multiplication of Fractions** The student chooses appropriately from a range of mental strategies to estimate answers and solve problems that involve the multiplication of fractions and decimals. The student can also use mental strategies to solve simple division problems with decimals. These strategies involve the partitioning of fractions and relating the parts to one, converting decimals to fractions and vice versa, and recognising the effect of number size on the answer, e.g., 3.6 × 0.75 = $\frac{3}{4}$ × 3.6 = 2.7 (conversion and commutativity); e.g., $\frac{2}{3} \times \frac{3}{4} =$ □, as $\frac{1}{3} \times \frac{1}{4} = \frac{1}{12}$ so $\frac{2}{3} \times \frac{1}{4} = \frac{2}{12}$ so $\frac{2}{3} \times \frac{3}{4} = \frac{6}{12} = \frac{1}{2}$ e.g., 7.2 ÷ 0.4 as 7.2 ÷ 0.8 = 9 so 7.2 ÷ 0.4 = 18 (doubling and halving with place value).	**Fractions, Ratios, and Proportions by Re-unitising** The student chooses appropriately from a broad range of mental strategies to estimate answers and solve problems involving fractions, proportions, and ratios. These strategies involve using common factors, re-unitising of fractions, decimals and percentages, and finding relationships between and within ratios and simple rates. e.g., 6:9 as □ :24, 6 × $1\frac{1}{2}$ = 9, □ × $1\frac{1}{2}$ = 24, □ = 16 (between unit multiplying); or 9 × $2\frac{2}{3}$ = 24, 6 × $2\frac{2}{3}$ = 16 (within unit multiplying) e.g., 65% of 24: 50% of 24 is 12, 10% of 24 is 2.4 so 5% is 1.2, 12 + 2.4 + 1.2 = 15.6 (partitioning percentages).

The Number Framework – Knowledge

Stage	Number Identification	Number Sequence and Order	Grouping/Place Value	Basic Facts	Written Recording
↑ Stage Zero: Emergent ↑ Stages One, Two, Three: Counting from One ↓	The student identifies: • all of the numbers in the range 0–10.	The student says: • the number word sequences, forwards and backwards, in the range 0–10 at least; • the number before and after a given number in the range 0–10. The student orders: • numbers in the range 0–10.	The student instantly recognises: • patterns to 5, including finger patterns.		
	The student identifies: • all of the numbers in the range 0–20.	The student says: • the number word sequences, forwards and backwards, in the range 0–20; • the number before and after a given number in the range 0–20; • the skip-counting sequences, forwards and backwards, in the range 0–20 for twos and fives. The student orders: • numbers in the range 0–20.	The student knows: • groupings within 5, e.g., 2 and 3, 4 and 1; • groupings with 5, e.g., 5 and 1, 5 and 2, ...; • groupings within 10, e.g., 5 and 5, 4 and 6, ... etc; The student instantly recognises: • patterns to 10 (doubles and 5-based), including finger patterns.	The student recalls: • addition and subtraction facts to five, e.g., 2 + 1, 3 + 2, 4 – 2, ... etc; • doubles to 10, e.g., 3 + 3, 4 + 4, ... etc.	The student records: • the results of counting and operations using symbols, pictures, and diagrams.

Stage	Number Identification	Number Sequence and Order	Grouping/Place Value	Basic Facts	Written Recording
Stage Four: Advanced Counting	The student identifies: • all of the numbers in the range 0–100; • symbols for halves, quarters, thirds, and fifths.	The student says: • the number word sequences, forwards and backwards, in the range 0–100; • the number before and after a given number in the range 0–100; • the skip-counting sequences, forwards and backwards, in the range 0–100 for twos, fives, and tens. The student orders: • numbers in the range 0–100.	The student knows: • groupings with 10, e.g., 10 and 2, 10 and 3, ... and the pattern of "-teens"; • groupings within 20, e.g., 12 and 8, 6 and 14; • the number of tens in decades, e.g., tens in 40, in 60.	The student recalls: • addition and subtraction facts to 10, e.g., 4 + 3, 6 + 2, 7 – 3, ...; • doubles to 20 and corresponding halves, e.g., 6 + 6, 7 + 7, $\frac{1}{2}$ of 14; • "ten and" facts, e.g., 10 + 4, 7 + 10 • multiples of 10 that add to 100, e.g., 30 + 70, 40 + 60.	The student records: • the results of mental addition and subtraction, using equations, e.g., 4 + 5 = 9, 8 – 3 = 5.
Stage Five: Early Additive	The student identifies: • all of the numbers in the range 0–1000; • symbols for the most common fractions, including at least halves, quarters, thirds, fifths, and tenths; • symbols for improper fractions, e.g., $\frac{5}{4}$.	The student says: • the number word sequences, forwards and backwards, by ones, tens, and hundreds in the range 0–1000; • the number 1, 10, 100 before and after a given number in the range 0–1000; • the skip-counting sequences, forwards and backwards, in the range 0–100 for twos, threes, fives, and tens. The student orders: • numbers in the range 0–1000; • fractions with like denominators, e.g., $\frac{1}{4}$, $\frac{2}{4}$, $\frac{3}{4}$, ... etc.	The student knows: • groupings within 100, e.g., 49 and 51 (particularly multiples of 5, e.g., 25 and 75); • groupings of two that are in numbers to 20, e.g., 8 groups of 2 in 17; • groupings of five in numbers to 50, e.g., 9 groups of 5 in 47; • groupings of ten that can be made from a three-digit number, e.g., tens in 763 is 76; • the number of hundreds in centuries and thousands, e.g., hundreds in 800 is 8 and in 4000 is 40. The student rounds: • three-digit whole numbers to the nearest 10 or 100 e.g., 561 rounded to the nearest 10 is 560 and to the nearest 100 is 600.	The student recalls: • addition facts to 20 and subtraction facts to 10, e.g., 7 + 5, 8 + 7, 9 – 6, ...; • multiplication facts for the 2, 5, and 10 times tables and the corresponding division facts; • multiples of 100 that add to 1000, e.g., 400 and 600, 300 and 700.	The student records: • the results of addition, subtraction, and multiplication calculations using equations, e.g., 35 + 24 = 59, 4 × 5 = 20, and diagrams, e.g., an empty number line.

Stage	Number Identification	Number Sequence and Order	Grouping/Place Value	Basic Facts	Written Recording
↑ Stage Six: Advanced Additive ↓	The student identifies: • all of the numbers in the range 0–1 000 000; • decimals to three places; • symbols for any fraction including tenths, hundredths, thousandths, and improper fractions.	The student says: • the whole number word sequences, forwards and backwards, by ones, tens, hundreds, and thousands in the range 0–1 000 000; • the number 1, 10, 100, 1000 before and after a given whole number in the range 0–1 000 000; • forwards and backwards word sequences for halves, quarters, thirds, fifths, and tenths, e.g., $\frac{1}{3}$, $\frac{2}{3}$, 1, $\frac{4}{3}$, $\frac{5}{3}$, etc. • the decimal number word sequences, forwards and backwards, in tenths and hundredths. The student orders: • whole numbers in the range 0–1 000 000; • unit fractions for halves, thirds, quarters, fifths, and tenths.	The student knows: • groupings within 1000, e.g., 240 and 760, 498 and 502, ...; • groupings of two, three, five, and ten that are in numbers to 100 and finds the resulting remainders, e.g., threes in 17 is 5 with 2 remainder, fives in 48 is 9 with 3 remainder. • groupings of 10 and 100 that can be made from a four-digit number, e.g., tens in 4562 is 456 with 2 remainder, hundreds in 7894 is 78 with 94 remainder. • tenths and hundredths in decimals to two places, e.g. tenths in 7.2 is 72, hundredths in 2.84 is 284. The student rounds: • whole numbers to the nearest 10, 100, or 1000. • decimals with up to two decimal places to the nearest whole number, e.g., rounds 6.49 to 6, rounds 19.91 to 20.	The student recalls: • addition and subtraction facts up to 20, e.g., 9 + 5, 13 – 7; • multiplication basic facts up to the 10 times tables (10 × 10) and some corresponding division facts; • multiplication basic facts with tens, hundreds and thousands, e.g., 10 × 100 = 1000, 100 × 100 = 10 000	The student: • records the results of calculations using addition, subtraction, multiplication, and division equations, e.g., 349 + 452 = 350 + 451 = 801, e.g., 45 ÷ 9 = 5, • demonstrates the calculation on a number line or with a diagram. The student performs: • column addition and subtraction with whole numbers of up to four digits. e.g., $\begin{array}{r} {\scriptstyle 11} \\ 476 \\ +\,285 \\ \hline 761 \end{array}$ $\begin{array}{r} {\scriptstyle 6991} \\ \cancel{7}000 \\ -\,586 \\ \hline 6414 \end{array}$

Stage	Number Identification	Number Sequence and Order	Grouping/Place Value	Basic Facts	Written Recording
↑ Stage Seven: Advanced Multiplicative		The student says: • the decimal word sequences, forwards and backwards, by thousandths, hundredths, tenths, ones, tens, etc.; • the number one-thousandth, one-hundredth, one-tenth, one, ten, etc. before and after any given whole number. The student orders: • decimals to three places, e.g., 6.25 and 6.3; • fractions, including halves, thirds, quarters, fifths, tenths.	The student knows: • the groupings of numbers to 10 that are in numbers to 100 and finds the resulting remainders, e.g., sixes in 38, nines in 68; • the groupings of ten, one hundred, and one thousand that can be made from a number of up to seven digits, e.g., tens in 47 562, hundreds in 782 894, thousands in 2 785 671; • equivalent fractions for halves, thirds, quarters, fifths, and tenths with denominators up to 100 and up to 1000, e.g., 1 in 4 is equivalent to 25 in 100 or 250 in 1000. The student rounds: • whole numbers and decimals with up to two places to the nearest whole number or $\frac{1}{10}$, e.g., rounds 6.49 to 6.5 (nearest tenth).	The student recalls: • division basic facts up to the 10 times tables, e.g., 72 ÷ 8. • fraction <—> decimal <—> percentage conversions for halves, thirds, quarters, fifths, and tenths, e.g., $\frac{3}{4} = 0.75 = 75\%$. The student knows: • divisibility rules for 2, 3, 5, 9, and 10, e.g., 471 is divisible by 3 since 4 + 7 + 1 = 12; • square numbers to 100 and the corresponding roots, e.g., $7^2 = 49$, $\sqrt{49} = 7$. The student identifies: • factors of numbers to 100, including prime numbers, e.g., factors of 36 = {1, 2, 3, 4, 6, 9, 12, 18, 36}; • common multiples of numbers to 10, e.g., 35, 70, 105, ... are common multiples of 5 and 7.	The student records: • the results of calculations using equations, e.g., 6 × 28 = 168, and diagrams, e.g., empty number line. The student performs: • column addition and subtraction for whole numbers; • short multiplication and division of a three-digit whole number by a single-digit number. e.g., $\overset{3\,2}{473} \times 8 = 3784$ $7\overline{)5488}$ → 784

Stage	Number Identification	Number Sequence and Order	Grouping/Place Value	Basic Facts	Written Recording
← Stage Eight: Advanced Proportional ·····		The student says: • the decimal word sequences, forwards and backwards, by thousandths, hundredths, tenths, ones, tens, etc., starting at any decimal number; • the number one-thousandth, one-hundredth, one-tenth, one, ten, etc. before and after any given decimal number. The student orders: • fractions, decimals, and percentages.	The student knows: • the number of tenths, hundredths, and one-thousandths that are in numbers of up to three decimal places, e.g., tenths in 45.6 is 456, hundredths in 9.03 is 903, thousandths in 8.502 is 8502; • what happens when a whole number or decimal is multiplied or divided by a power of 10, e.g., 4.5×100; $67.3 \div 10$. The student rounds: • decimals to the nearest 100, 10, 1, $\frac{1}{10}$, or $\frac{1}{100}$, e.g., rounding 5234 to the nearest 100 gives 5200.	The student recalls: • fraction ⟷ decimal ⟷ percentage conversions for given fractions and decimals, e.g., $\frac{9}{8} = 1.125 = 112.5\%$. The student knows: • divisibility rules for 2, 3, 4, 5, 6, 8, and 10, e.g., 5 632 is divisible by 8 since 632 is divisible by 8, e.g., 756 is divisible by 3 and 9 as its digital root is 9; • simple powers of numbers to 10, e.g., $2^4 = 16$, $5^3 = 125$ The student identifies: • common factors of numbers to 100, including the highest common factor, e.g., common factors of 48 and 64 = {1, 2, 4, 8, 16}; • least common multiples of numbers to 10, e.g., 24 is the least common multiple of 6 and 8.	The student records: • the results of calculations using equations, e.g., $\frac{3}{4} \times 28 = 21$, and diagrams, e.g., double number line. The student performs: • column addition and subtraction for whole numbers, and decimals to three places; • short multiplication and division of whole numbers and decimals by single-digit numbers, e.g., $4\overline{)58.60}$ (14.65) • multiplication of three- or four-digit whole numbers by two-digit whole numbers, e.g., 763 × 49 = 6867 + 30520 = 37387

Victoria
UNIVERSITY OF WELLINGTON
Te Whare Wānanga
o te Ūpoko o te Ika a Māui

LIBRARY

VICTORIA UNIVERSITY OF WELLINGTON LIBRARY
3 7212 01445574 3