CARRYING OUT A PRACTICAL INVESTIGATION: THE STUDENT-EXPERIENCED CURRICULUM

Anne Hume

Thesis submitted in partial fulfilment

of the requirements for the degree of

Doctor of Education

University of Waikato

2006

ABSTRACT

Assessment of student learning for qualifications has long been contentious, particularly the reliability and validity of various assessment methods and the impact of assessment on the nature of student learning. There is, for example, controversy about norm-referencing versus standards-based assessment of student learning with different stakeholders holding different political agenda. In this study these issues are investigated by research into a recently implemented qualification for secondary school students in New Zealand – the National Certificate of Educational Achievement (NCEA). Proponents of NCEA suggest this regime acknowledges a wider range of student achievements than previous qualifications. Opponents of NCEA lament the ‘fragmented’, ‘atomistic’ approach to learning, and the lack of incentive for students to compete and strive for excellence. While other New Zealand research reports on national trends in the way school programmes are responding to the changes associated with the implementation of NCEA, little is known about the classroom curriculum students are actually achieving.

This interpretive study presents findings from classroom-based case studies investigating the nature of the student-experienced curriculum for the NCEA Science Achievement Standard 1.1 *Carrying out a practical investigation with direction.* These findings indicate that the achievement standard is exerting a strong influence over the nature of student learning, with students experiencing purposeful and focused learning within a structured teaching programme. The emphasis in the standard on fair testing has implications for students’ understanding of the nature of scientific inquiry by limiting their exposure to the range of methods that scientists use in practice. Student learning tends to be mechanistic and superficial rather than creative, critical and life-long.

The findings have important implications for writers of national curriculum and assessment policy and for teachers of science. Suggestions are offered for ways of enhancing the learning of science under the NCEA qualification.

ACKNOWLEDGEMENTS

I would first like to extend my sincere gratitude to my chief supervisor, Richard Coll, for his skilled and good-humoured guidance as I worked towards completion of my doctoral thesis. His prompt and painstaking feedback made this task, arduous at times, a rewarding learning experience for me. I especially appreciated Richard's generosity in terms of his time and willingness to accommodate my rather disjointed research timetable into his own hectic working schedule. My thanks also to Bronwen Cowie, my second supervisor, for her timely big picture comments that helped to put things into perspective and drew me back to the key issues of my thesis.

I am indebted to the participating teachers and students in my study who gave so freely of your time and thoughts. Your readiness to share your teaching and learning experiences honestly and candidly with me provided a wealth of data and invaluable insights to inform my research. Good luck to all of you in your future endeavours.

Special thanks go to my former colleagues at the Education Review Office (Tainui) for their encouragement and assistance during the early stages of the doctorate. I am most grateful to Ian Hill for his support of my doctoral studies, in particular the all-important gift of time at a critical stage in my data gathering phase. Thanks also to my review team members who covered for me while I juggled my roles of reviewer and researcher.

On the home front my husband Terry has been a stoic and loyal supporter of my research efforts. A big vote of thanks for your wonderful evening meals and for taking up golf – that helped to ease my feelings of guilt for all those long hours you were left on your own! My research assistants Lewis and Floyd provided me with much needed companionship and diversion during my long hours at the computer.

Lastly I'd like to dedicate this thesis to my late mother Marjorie Nordgren. I know she would have so enjoyed coming to my capping ceremony!

TABLE OF CONTENTS

CHAPTER FOUR

CHAPTER FIVE

LIST OF TABLES

CHAPTER ONE

CHAPTER ONE

INTRODUCTION AND BACKGROUND

Overview

This chapter provides the rationale for my research into the science curriculum experienced by New Zealand students as they work towards their National Certificate of Educational Achievement (NCEA) qualification. I background my personal involvement in science education, and outline past and current developments in curriculum, assessment and qualifications within the New Zealand context that have implications for teaching and student learning. These implications are signalled and discussed in relation to the possible effect they may have on the student-experienced curriculum in science. Justification for my investigation is supported by the paucity of classroom-based research available to inform ongoing developments in curriculum, assessment and qualifications, and effective teaching and learning practice. I identify key gaps in the current knowledge to justify the focus of my investigation and the design of my research questions.

1.1 The Researcher's Personal Background and Involvement In Science Education

Until embarking on this study four years ago, my experience with formal research had been limited to completing a directed research project in science education 12 years before as part of a Master of Education degree. In that project I had investigated the feasibility of negotiating curriculum through children's questions (Hume, 1991), and this work introduced me to curriculum design processes. During the following years, while teaching science to secondary school students and facilitating teacher professional development in science education, I became increasingly aware of two things. Firstly, my ability to teach scientific investigation work effectively was limited because my own experience and understanding of scientific research was actually quite superficial, despite having an undergraduate degree in science and a post-graduate qualification in science education. Secondly, I was experiencing a growing desire to carry out my own education research with more rigour and in greater depth. I wanted to learn more about what really constituted quality scientific investigation in the classroom and how best to teach it.

This awareness of my own lack of depth in education research was heightened when I entered the field of education evaluation, where I was required as a Government reviewer for the Education Review Office (ERO) to make professional judgements about the quality of teaching and learning in schools and classrooms. Evaluation, I soon discovered, shares many common features with research (Cohen, Manion & Morrison, 2000), particularly the theoretical frameworks that guide these forms of human activity; the systematic and planned collection, analysis and interpretation of information; and the methods and tools used in data-gathering and handling. Where the two activities could be said to differ is in their purpose – research is primarily exploratory and investigative, intent on generating, legitimating and advancing knowledge (Cohen et al., 2000; Lather, 1992; University of Waikato, 2001) while evaluation provides information to assist decision-makers in their judgements (Keeves, 1998), often of the worth of "given initiatives" (Cohen et al., 2000, p. 38). However, even this difference is becoming increasingly less obvious, since many evaluative studies in education are now also contributing to a growing understanding of educative processes and systems (Keeves, 1998), particularly from the perspectives of participants in new approaches to evaluation such as *responsive evaluation* (Guba & Lincoln, 1989). If I was to become an effective evaluator a more authentic understanding of research on my part was becoming vital, and I formed the view that the necessary knowledge, skills and capabilities in research could only be achieved by my personal engagement in a project that would give me genuine and extensive experience in research processes. I accepted the challenge and decided to undertake a doctoral study in an area of special interest to me – the teaching and learning of scientific investigation.

My research interest in science education stems from 25 years experience in science teaching, over 30 years married to a scientist, and close involvement for over 15 years in many stages of the development, writing, and implementation of the *Science in the New Zealand Curriculum* (SiNZC) (Ministry of Education [MoE], 1993a). My involvement with the SiNZC included acting as a: contributing writer to the document *Draft Forms 1-5 Science Syllabus for Schools* (MoE, 1990); participant in the Learning in Science Project (Teacher Development); member of the writing team for the document SiNZC; facilitator for the MoE science contract for professional development of teachers; head of a secondary school science department with responsibility for implementing the science curriculum; teacher of science; national examiner and moderator for science; writer of textbooks to support the new science curriculum; parent of students learning under the new science curriculum; and reviewer of school science programmes. I was also member of the expert panel in science for the National Certificate of Educational Achievement (NCEA), a New Zealand qualification for senior students based in large part on national curriculum statements. Through these roles and experiences I had become conscious of the many factors that influence classroom teaching and learning in science, and how the impact of these factors vary from one classroom to the next. Curriculum and qualification requirements, teaching approaches, classroom dynamics, school culture and the physical environment were just some of the many factors affecting classroom teaching and learning. What I began to sense, particularly as an evaluator required to judge the quality of education provided in New Zealand schools, was the growing realisation in the educational community of how little was known about what students were actually experiencing and learning in classrooms. This realisation is now manifested in a strong emphasis from government on student achievement (MoE, 2004b) and in identifying ways to measure and improve learning. Thus the reality of the student-experienced curriculum in science has emerged as a potentially worthwhile research topic for me to explore, both from the perspective of a personal research interest and that of a national goal in education.

In my own science teaching I had put a great deal of time and thought into trying to create teaching and learning episodes that allowed my students to engage in more authentic, contextualised scientific inquiry such as practicing scientists might do (Brown, Collins & Duguid, 1989; Woolnough & Allsop, 1985). This redesign of my teaching programme was in response to new curriculum and assessment directions that I was learning about through personal involvement in national curriculum and qualification development, and through their implementation in my science department and classroom. I was familiar with many of the issues associated with the teaching and learning of practical scientific investigation, and aware that under recent qualifications reforms greater attention was being paid to science processes. There were likely to be ramifications for classroom practice from these reforms, particularly in the senior secondary area of schooling where practical skills were to be internally assessed. In my view, it seemed research into practical scientific investigation in senior secondary classes had great potential for findings that would inform the wider educational scene.

The following sections in this chapter introduce key educational developments in New Zealand and the wider scene that have a direct bearing on both present and future senior science classroom programmes in New Zealand. The issues and research opportunities that emerge as a result of these influences are signalled and subsequently developed into research questions for this thesis.

1.2 The Status of the Current National Science Curriculum

The SiNZC has been the national curriculum policy for schools in New Zealand for over 10 years. The development of this curriculum, including its implementation in schools, represented a huge investment of time, human and financial resources and was not without controversy (Bell, Jones & Carr, 1995). In 2002 a ‘curriculum stock-take’ was commissioned by the government to analyse New Zealand curriculum reform of the past 10 years. This analysis (MoE, 2002a) examined the appropriateness of the overall curriculum in terms of the current educational, social and economic climate; the purposes of the curriculum; and the quality of the curriculum in contributing to improved student outcomes, meeting the expectations of a range of stakeholders, and against comparable international curricula. Among the recommendations of the stock-take report was retention of frameworks for the essential learning areas similar in structure to the existing frameworks of the New

Zealand Curriculum (NZC), but modified to include review, refinement and clarification of the learning outcomes. In response to the recommendations of the stock-take the New Zealand Curriculum Project (NZCP) was set up in 2003 to build on, and the first stage of this new project was an official review of the curriculum (MoE, 2003a). Contributions were sought both from within and outside the New Zealand educational community, and consultation over developments with principals and teachers and other stakeholders continued from late 2003 into 2005. This co-construction aims to ensure that curriculum development is guided by research findings, teachers' experience, and school realities. Draft curriculum statements were released in 2005 for comment and the revised curriculum is due to be published in 2007.

In reports on progress to date (MoE, 2004a; 2005) the NZCP forecasts that by 2008 or 2009 all current curriculum statements, including that for science, will be replaced by one overarching new document, similar in look and layout to the current New Zealand Curriculum Framework (MoE, 1993b). The project aims to reframe, refocus and revitalise the current curriculum. The new single document will contain essence statements encapsulating the fundamental ideas for each learning area, accompanied by a reduced number of refined achievement objectives. A new cluster of key competencies focusing on the skills, knowledge, attitudes and values used together by people across a variety of life contexts are being developed to replace the existing *essential skills and attitudes*. These competencies are to be integrated into the new achievement objectives. There will be a focus on effective teaching and strengthening school ownership of curriculum.

1.3 Curriculum Review

Review and implementation of a new curriculum is costly and time-consuming. To avoid previous mistakes and gain insights into these processes curriculum redevelopers can learn from examining past experience. A number of studies and critiques of curriculum development and implementation in the New Zealand context do exist from which valuable lessons might be learned (Austin, 1999; Bell & Baker, 1997; Bell et al., 1995; Cowie, Jones & Harlow, 2003; Haigh, 1995; Lewthawite, 1999; Lewthawite, Stableford & Fisher, 2001; Loveless & Barker, 2000). Examination of the early evaluations done of the NZC, including the SiNZC, reveal that most were initially confined to the effectiveness of the professional development programmes for teachers, which supported the implementation of the curriculum (Dewar & Bennie, 1996; Gilmore, 1994; Poskitt, 1992). These evaluations concentrated on the outcomes for teachers. Remarkably, there was little mention or attention given to the effect on student learning or change in student achievement (Gilmore, 1994; Thornton, 2003). Teachers' opinions were frequently sought in such studies, but there were no observations of teachers' classroom practice after the professional development by those carrying out the evaluations. More recently as part of the NZCP, the National School Sampling Study: Science (NSSS) sought teachers' views again on their experiences implementing the SiNZC, and on the factors that had constrained and enabled implementation (Cowie et al., 2003; MoE, 2003b). As in earlier evaluations, students' views were not canvassed, and few findings emerged about student achievement. Thus it appears that research done to date in New Zealand on the science curriculum implementation sheds little light on the actual nature of the classroom curriculum experienced by students.

This lack of focus on outcomes for students in curriculum implementation is also reflected in the wider educational field, where research into students' actions, perceptions and opinions about the nature of their learning tends to be sparse. In fact student voices or perspectives on schooling, learning and assessment are noticeably absent in the literature generally (Cowie & Bell, 1999; Leach & Scott, 2003; Linsell, 1999; Powell & Anderson, 2002; Reay & Wiliam, 1999). Many authors now support the contention that "a focus on teaching is not equivalent to a focus on learning" (Wenger, 1998, p. 266). To evaluate the real worth of any curriculum there is a need to concentrate on the learning outcomes for students (Bell & Baker, 1997; Burt & Davidson, 1998; ERO, 2000). Research that can identify current outcomes for students, and features of the student experienced curriculum that are linked to improved student learning, has the potential to inform the curriculum redevelopment process and assist future implementation.

1.4 Teachers' Response to the SiNZC

Teacher response to the SiNZC has been generally favourable (Baker, 1999; Bell et al., 1995; MoE, 2003b). Teachers were involved in the writing of the document and the mix of theoretical and philosophical views on science education in the document accommodates most teachers' views. Hipkins and Barker (2002, p. 10) suggest that in fact many teachers simply interpreted the document as "business as usual" and continued their classroom practice with little change. Some teachers, under the pressure of content coverage and high stakes national exams, coped by simply incorporating the new, much broader content into what they traditionally taught (English, 1995). Kennedy (1997) speaks of a similar reaction by teachers, a kind of “defensive rigidity” (p. 8), to mandated change to curriculum in the USA, and this trend is still evident in more recent reports on curriculum implementation in the USA (Powell & Anderson, 2002). Practitioners who are faced with what they perceive to be radical changes tend to rely on familiar and established routines for ‘survival’ (Atkin & Black, 2003). Often their capacity for change is overestimated by policy-makers.

To take account of outcomes for students, one means of judging the effectiveness of a national curriculum development and its implementation would be to gauge the degree of match between the ‘intended curriculum’ and the ‘operational curriculum’ (McGee, 1997). The term ‘intended curriculum’ can include curricula depicted in government policy, school documents such as programme schemes, or by teachers in their planning and delivery. The operational curriculum is that actually experienced and learned by the students in classrooms. McGee argues the closer the match between the intended curriculum as articulated in national policy and the operational policy, the more effective the curriculum development and implementation process has been. The absence of research data about such translation and/or transformation of national curriculum policy at school and classroom levels is identified in the recent curriculum stock-take (MoE, 2002a). Bell and Baker (1997) maintain, however, that a curriculum and its implementation can only be judged effective if the student-experienced curriculum changes in a way that improves learning. This view of curriculum effectiveness necessitates a clear understanding of what comprises improved learning. It needs to be made clear whether this is improved learning of the intended curriculum, or of learning judged worthwhile by others concerned with education that may or may not be part of the intended curriculum. This thesis aims to use these different measures of curriculum effectiveness as a means of exploring the nature of the student-experienced curriculum, and the influence various factors have on that experience.

1.5 From Intended Curriculum to Operational Curriculum

In actuality moving from policy document to the operational curriculum is not necessarily a rational and linear process (Atkin & Black, 2003; McGee & Penlington, 2001; MoE, 2002a), and it is more realistic to view national curriculum policy as “the start of a cascade of interpreted curricula” (Carr et al. 2001, p. 18) that influence what students experience and learn in the classroom. The process of implementing curriculum policy can be seen as one of interplay between what a curriculum statement says and the various interpretations and emphases afforded it by supporting materials, agencies, schools and teachers (Knapp, 2002; Spillane, 2004). In senior secondary schooling the impact of assessment for national qualifications is also a major consideration (English, 1997). These influences largely determine the nature of the teacher’s intended curriculum, but there is another multi-layered and multi-dimensional set of processes to consider before eventually arriving at the student-experienced curriculum. Teachers’ personal beliefs and values, their knowledge and skills in subject content and pedagogy, and ability to adapt to change all impinge on the nature of the curriculum they deliver in their classrooms (Hargreaves & Fullan, 1992). In turn, students’ prior knowledge, ability to make connections and access meaning through language, relationships with teacher and peers, feelings of self-esteem and motivation, and classroom social practices and interactions all have filtering and modifying effects on the intended curriculum. In sum, the whole of the cognitive, social and language processes occurring in the classroom plays an important role in shaping the actual curriculum received by students because these processes determine how students think and learn (Nuthall, 1997).

1.6 The Impact of Assessment for Qualifications on Student–experienced Curricula

Assessment for qualifications is a key area to explore in this study in terms of the impact this practice has on the student-experienced curriculum. The National Certificate of Educational Achievement (NCEA) qualification is a manifestation of reforms first introduced in the National Qualifications Framework (NQF) in 1991 by the New Zealand Qualifications Authority [NZQA] (Lee & Lee, 2001). These reforms heralded a new assessment culture in New Zealand (Lennox, 1995) where assumptions underlying the existing national assessment and qualifications procedures were challenged. NCEA was introduced in 1996 alongside other existing qualifications on the NQF (Lee & Lee, 2001) and it was based on components known as *unit standards*. Unit standards are statements, developed from national curriculum statements and/or existing examination prescriptions (Hipkins & Vaughan, 2002), which describe in the form of specific performance criteria what students need to know and do in order to gain credit. Judgement of student achievement occurs at two levels: credit and non-credit. In response to recommendations made in a review of the NQF (MoE, 1997) NCEA took on greater importance and as a result is to subsume other qualifications as the principal school leaving qualification for New Zealand students. Another component, the *achievement standard* was introduced as the essential building block of NCEA in the academic subjects of the school sector. Achievement standards judge student performance at four levels (non achievement,

achievement, achievement with merit and achievement with excellence), but use less specific performance criteria than those used for unit standards

The first stage of the NCEA (Level 1) implementation occurred in 2002 at the Year 11 level of secondary schooling. Today students can accumulate credits for the NCEA qualification at Levels 1, 2 and 3, by demonstrating they have met or exceeded the predefined outcomes of the achievement and/or unit standards (NZQA, 2001). As noted earlier, achievement standards allow judgement of student achievement at three levels: ‘achieved’, ‘achieved with merit’ and ‘achieved with excellence’. These descriptors are intended to give parents, tertiary institutions and employers a more detailed picture of overall student achievement (NZQA, 2001). While the vocationally-related unit standards generally need internal assessment, at least half the achievement standards for conventional school subjects are externally assessed, to address concerns about issues to do with internal assessment, such as moderation and teacher workload (Lee & Lee, 2001).

Qualifications like NCEA are a form of quality assurance for national educational outcomes (Black, 2001; Lee & Lee, 2001; MoE, 1997), and assessment for qualifications is considered ‘high stakes’ to all involved. Such assessment will often drive senior school and classroom programmes because of teachers’ desire to secure good results for their students (Black, 2001; 2003a). In a report to the Ministry of Education on proposals for development of NCEA, Black (2003a) makes the point that NCEA places new assessment demands on teachers. These demands will in turn affect classroom teaching and learning practice, as teachers explore and exploit the means of securing maximum success for their students. Black calls for continual evaluation and review of the evolving roles of teachers, teaching and learning in classrooms and assessment practice, to inform the new qualification system as it stabilises and matures.

Recent calls for input to the curriculum review by the MoE (2003a), and the continued development of NCEA (Black 2003), provide opportunities for valuable periphery-to-centre information flow (Bell & Baker, 1997). A research project into the current nature of the student-experienced curriculum in the senior secondary area, particularly under an evolving qualification regime, thus seems highly appropriate and timely.

1.7 Assessment for Learning (Formative Assessment)

According to many authors assessment for student learning is now playing a key role in educational reform (Bell & Cowie, 2001; Black & Wiliam, 1998; Crooks, 2002; Murphy, 2001; Tytler, 2003). There is a growing belief that to support and strengthen learning there needs to be a match between assessment practice and current understanding about the nature of learners and of the learning process (Hodson, 1996; Murphy, 1995; Powell & Anderson, 2002). Research points to classroom-based formative assessment as a powerful means of improving student learning (Bell & Cowie, 2001; Black & Wiliam, 1998; Clarke, 2001; MoE, 2003; Moreland, Jones & Chambers, 2001).

Traditionally, summative assessment methods provide data to inform qualification decisions, and arguably formative assessment should play a significant part in the learning that is part of acquiring qualifications. However, opinion is divided on whether formative assessment can sit comfortably within a qualification structure where classroom teachers are required to make final creditation judgements (Black, 2003; Elley, 2003). Some authors like Harlen (1998) believe tensions between summative and formative functions may threaten the validity, reliability, and fairness of each assessment method and undermine the usefulness of both sets of assessment information. However, Wiliam (2000) a proponent of formative assessment use within qualifications, points out a separate system reliant on external agencies for summative assessments excludes teachers from the assessment of their students, and limits the amount of valid data that can be gathered to a narrow range of students' achievements. Expecting teachers to make summative decisions without divulging formative information as teaching and learning progresses would be detrimental to their student’s continued learning. The role of formative assessment within a qualification such as NCEA is thus far from clear, and warrants investigation.

1.8 Science Investigations as a Context for the Inquiry

Two recent studies undertaken by the New Zealand Council for Educational Research (Hipkins & Vaughan, 2002; Hipkins et al., 2004) indicate that, despite the flexibility of the new NCEA qualification, schools still deliver science to most students in ‘traditional-discipline’ options for both Years 11 and 12. These courses continue to reflect the traditional ways of thinking in education about the structure of science, and are mainly assessed by achievement standards. Findings, from the investigation of six case study schools (Hipkins et al., 2004), probably reflect the development by NZQA of a subject-specific suite of achievement standards for each year level in science. The schools in these case studies have used these specific standards as the basis of the assessment of their Year 11 and 12 science options for NCEA, and in turn the school curriculums tends to be organised around the divisions dictated by the separate standards. The science suite is made up of seven achievement standards, five of which are externally assessed by national examinations, and classroom teachers assess the other two internally (Hipkins et al., 2004).

It is the forum of internal assessment for NCEA that offers the greatest potential for my research to inform current practice in assessment for learning, because it is here that some of the issues related to the viability of effective formative assessment are most likely to emerge (Harlen, 1998; Reay & Wiliam, 1999). At Year 11 early indications are that the most commonly chosen of the internal science standards programmes in schools is the internally assessed Science Achievement Standard 1.1 *Carrying out a practical investigation with direction* (Hipkins & Vaughan, 2002; Hipkins et al., 2004). Not surprisingly, students rated interesting and practical learning activities high on the list for enjoyment of learning in science for NCEA (Hipkins et al. 2004, p. 186). From the preceding descriptions of the current New Zealand education scene it can be seen that the specific teaching and learning classroom context of Science Achievement Standard 1.1 offers scope to investigate the impact of a wide range of influences on the student-experienced curriculum in scientific inquiry. These influences include:

- teachers' personal beliefs and values about teaching and learning
- teachers' knowledge and understanding of the nature of scientific inquiry, and the curriculum interpretations they deliver in classrooms
- pedagogical approaches and strategies
- teacher- student relationships and other social processes operating in the classroom such as language
- interests and learning needs of students

- new assessment demands on teachers and students, and
- the formative-summative tensions inherent in internally assessed components of a qualification.

1.9 Research Questions for this Thesis

To operationalise the purpose of this investigation and gain access into the field my main research aim, of investigating the science curriculum experienced by students as they work towards their NCEA qualification, can be addressed by seeking answers to these key questions:

- *What* science are New Zealand science students learning in NCEA classroom programmes for Science Achievement Standard 1.1 *Carrying out a practical investigation in science with direction*?
- *Why* and *how* are New Zealand science students learning the science they learn in NCEA classroom programmes for Science Achievement Standard 1.1?
- *What* match is there between the intended science curricula (those of the SiNZC and the teacher) and the operational science curricula for New Zealand science students in scientific investigation within the assessment of NCEA Science Achievement Standard 1.1 *Carrying out a practical investigation in science with direction*?

By revealing the learning reality of students I hope to highlight influences and aspects of classroom life that impact on the nature of their science learning as they work towards the partial fulfilment of a national qualification. *What*, *why* and *how* they learn, and the degree of match between the intended and operational curriculum should generate findings that have implications for curriculum redevelopment and pedagogy that seek to improve student learning.

1.10 Summary

This first chapter has provided the background to this study, including my personal research interest in this topic, the underlying reasons based on the research literature and events in the New Zealand education scene that call for such a study, and the research questions. The next chapter reviews findings from the research literature that are relevant to the research questions.

CHAPTER TWO

CHAPTER TWO

LITERATURE REVIEW

Overview

This chapter is a critical analysis of findings from the existing literature pertinent to the research questions in this study. This critique of current evidence and thinking, drawn from the fields of curriculum and learning theories, pedagogy, assessment and qualifications, and science education then informs the research methodology, and provides valuable reference points for the later analysis and interpretation of the research findings.

This review identifies and explores the multifaceted and interdependent relationships that exist between curriculum design, pedagogy and assessment. Current thinking on how these processes interact to affect learning is examined, along with the impact such thinking may exert on curriculum development and implementation internationally. A look at these trends in the curriculum field of science and investigatory skills follows, and the review concludes with an account of such developments in the current New Zealand science education scene.

2.1 Introduction

This section introduces key concepts and discussion points that will be further elaborated upon during the literature review, and signals how arguments in the review will develop.

'Curriculum' is a concept underpinning all the research questions in this study, and discussion of its meaning forms the focus of the early sections of this chapter. This exercise reveals subtle differences in meaning for 'curriculum' depending on the context in which the term is used, or the purpose and perspective of the user. Despite these shades of meaning all definitions of curriculum inherently involve an element of decision-making and choice about what is considered worthwhile learning (Schubert, 1986) by participants and interested parties. From national level as policy-making (the intended curriculum), to school and classroom level as implemented policy (the implemented or enacted curriculum), to that experienced and attained by students (the operational curriculum), curriculum is contestable and requires groups and/or individuals to make judgements about learning (Carr et al., 2001). The operational curriculum can be viewed as the culmination of decisions various stakeholders, at different levels or 'sites' within an education system, make about worthwhile learning (English, 1997). Its nature is shaped and determined by many layers of interpretation including those of government and its agencies, publishers of resource materials, school communities and management, teachers and students themselves (Carr et al., 2001; Knapp, 2002). This review attempts to explain how these layers of curriculum interpretation occur by exploring the influence political ideologies, learning and pedagogical theories, assessment of learning including qualifications, teachers' personal beliefs and values about teaching and learning and students' learning dispositions and attitudes have on curriculum design and implementation.

What students ultimately come to experience and learn of the curriculum delivered in classrooms is referred in the literature by the synonymous terms 'operational', 'student-experienced' and 'attained' curriculum. Teachers and students are central to the learning of this curriculum, where the teacher could be said to play an intervening role and students a mediating role in reaching their respective interpretations of the curriculum. Research suggests teaching and learning processes in classrooms are at the centre of complex and dynamic interactions between curriculum, pedagogy and assessment (Carr et al., 2001). It seems clear that the pedagogical and assessment approaches teachers use need to be aligned with the intent of the official curriculum, if there is to be a close match between the intended curriculum and the operational curriculum (McGee & Penlington, 2000). However, once designed and delivered as the implemented curriculum by teachers, the curriculum students come to experience and learn is not necessarily what the teacher intended or foresaw. Students mediate what is taught and learned in the classroom, creating their own interpretation of the teacher's intended curriculum (Nuthall, 1997). Their existing understandings, experiences and dispositions have considerable impact on the way students respond to this enacted curriculum, and what they consequently think and learn. Also students use their personal awareness of how social processes in classroom culture operate to reconcile their learning with influences and perceptions impacting on them as individuals. Feedback from others and feelings of self esteem feature prominently in the nature of the achieved curriculum. The important influence assessment has on the nature of the student-experienced curriculum is highlighted, particularly the effects teachers' formative and summative assessment practice can have on student learning for a qualification. Findings from the literature review indicate that the total sum of all processes and activities going on in a classroom influences the nature of student thinking and learning occurring.

The review moves on to consider the issues of curriculum, pedagogy and assessment as they relate to the current teaching and learning of scientific inquiry in the context of science education internationally. The content and abilities deemed important by science educators for students to acquire are pinpointed, accompanied by commonly held rationales for their inclusion. Pedagogical and assessment strategies needed to promote these desired learning outcomes are introduced and discussed, with the engagement of students in independent authentic inquiry and holistic contextualised assessment featuring prominently as desirable practice. However, there is recognition within the literature that a close match between the authentic practice of science and that experienced by students in education will always be problematic for teachers because evoking the complex and various ways in which scientists perform investigations within existing or favoured pedagogies and school contexts is a difficult task. Simplified portrayals of scientists' work typically result. Similarly, holistic assessment of authentic student inquiry could prove to be difficult for teachers since this practice requires deep and extensive understanding of the nature of scientific investigation. Findings from this literature indicate that many science teachers do not possess in-depth understanding or appreciation of how science really works.

The review then turns to the New Zealand educational scene, in the context of science, to give an account of how national curriculum and qualification development and implementation has occurred, including teacher development, and the impact to date of these developments on classroom curricula. In this account the political and educational background to the *Science in the New Zealand Curriculum* (SiNZC) is covered, including the key ideologies influencing in its content. The structure and content of this policy document is described and discussed, paying particular attention to the treatment of scientific investigation. The lead-up and development of the newly introduced New Zealand qualification, the National Certificate of Educational Achievement (NCEA), is outlined together with its structure and underpinning philosophy. The potential for summative–formative assessment tensions within such a qualification are identified and evaluated.

Finally, the scene is set for investigating the nature of the student-experienced curriculum in scientific inquiry under SiNZC and NCEA, in the context of a particular standard. This standard, called Science Achievement Standard 1.1 *Carry out a practical science investigation with direction,* is the most commonly taught qualification standard in scientific investigation for the first level of NCEA. It provides a context that is representative of many Year 11 secondary students' experiences in learning about scientific inquiry in New Zealand.

2.2 Exploring the Concept of Curriculum

As a concept curriculum has variations in meaning dependent on context and usage (MoE, 1993a) that can be teased out to reveal multiple interconnected layers (McGee & Penlington, 2001a). Teachers, for example, may view the "curriculum" as a policy document, school managers as the teaching and learning programmes going on in their school, and tertiary educators as a means of defining approaches to teaching and learning. Therefore, in any discussion about curriculum it is important to be aware of the different interpretations that can be assigned to the term, and to take care that its intended meaning in the text is clearly communicated to the reader. Key interpretations of the term ‘curriculum’ from the literature are established in the next two sections for use in this thesis, and some of the forms of curriculum defined.

2.2.1 Curriculum as a national policy statement

As a national policy statement curriculum owes its nature to the political context in which it was developed (Atkin & Black, 2003; Carr et al., 2001; Kennedy, 1997; McGee, 1997; Rudolph, 2002). Curriculum in this form is an ideological selection of what is deemed worthwhile knowledge (Schubert, 1986) from a range of knowledge, and is very much an exercise in political decision-making. In their review and critique of the development of the New Zealand science curriculum Bell et al. (1995) indicate that many different groups or stakeholders seek to have an influence on the outcome of such a national curriculum development process. As Carr et al. (2001, p. 18) point out: “The process is political involving contestation and debate”. Stakeholders may have educational, social, political or historical interests in the curriculum, with each group contributing towards a set of selection criteria for curriculum content (Rudolph, 2002). This selection process inevitably results in a specified curriculum (Murphy, 2003) that is a partial or selective representation of the nature of the knowledge domain, and claims are made about the potential of the domain for students’ learning. The selected content gains credibility in terms of what ‘counts’ as worthwhile knowledge, and what does not, by schools, teachers and communities when it is disseminated and implemented across schools. Orpwood (2001) makes the point that a national curriculum, for subjects such as science, should be constantly undergoing review as the knowledge domain itself evolves, and curriculum designers find new answers to the question of what is considered to be knowledge of worth.

These National policy curriculum statements commonly portray not only selected content, but also views on theories of learning and teaching. Whether expressed explicitly or implicitly these perspectives tend to reflect what McGee and Penlington (2000, p. 3) identify as the "synergistic relationship" between the paradigms of learning theory and those of teaching. In other words, curriculum writers have come to recognise that the combination of a learning theory with a complementary pedagogical approach is likely to have a more successful outcome educationally than treating learning theory and teaching in isolation. Thus a curriculum promoting a behaviourist theory of learning, where it is believed learners gradually acquire knowledge through the strengthening of association between a stimulus and a response, is likely to develop and promote a transmissive model of teaching. Such teaching would involve an instruction process where knowledge is transferred from teacher to student (Neyland, 1995), and learning promoted via techniques such as drill and practice.

2.2.2 Curriculum as a process of implementation

Curriculum in a broader sense of the term can take on a dynamic dimension and involve more than just national statements in the official curriculum (Bell & Baker, 1997). Curriculum can also include its implementation, as "the process of transforming the *intended curriculum* into the *operational curriculum"* (McGee, 1997, p. 15, added emphasis). This process necessitates the translation of national statements into school schemes and lesson plans, the selection and design of appropriate teaching approaches and learning activities, and student learning and assessment of that learning. This view of implementation has been described as a ‘centre-periphery’ model (McGee, 1997), where curriculum philosophy and direction is initiated and defined at the centre, and then disseminated to schools and teachers.

Others do not see the implementation process following a simple linear model of development (Atkins & Black, 2003), but rather one involving interplay between various influences that ultimately impacts on the classroom curriculum. In this model the curriculum experienced and learned by students in classrooms is the result of the interplay of different sites of influence or promotion within an education system. A site as defined by English (1997) is a context or ‘arena of action’ where participants have shared understandings of concepts and ideas due to the shared social contexts. Such sites in an educational system could be national curriculum documents, professional development provision, instructional texts, national qualifications, and classroom programme planning and delivery (Crooks, 2002b; English, 1997; McGee & Penlington, 2001a; Orpwood, 2001). This ‘landscape’ view serves to introduce the wider range of individual people and groups (Lewthwaite, Stableford & Fisher, 2001) who have vested interests and influence on the curriculum received by students. The range includes not only politicians and subject experts, but also parents, boards of trustees (i.e., the governing body of a school), curriculum advisers and facilitators, teacher developers, principals, school curriculum leaders, education evaluators, school curriculum development teams, other students, and of course, the classroom teacher. All bring, to varying degrees, their views on what should be learned and how, to the curriculum decision-making process, and because of this Black (2001) describes curriculum implementation as a form of 'social engineering'. The interpretations these stakeholders bring to the curriculum are made in light of individuals' experiences, belief systems and values, and there may or may not be consistency of understanding or commitment (Orpwood, 2001). At all levels curricular decisions are made that are contestable, and affected by a range of ideological, political, economic, philosophical, social or cultural influences (Carr et al., 2001). According to McGee (1997) curriculum development, including implementation, is a value-laden activity. If implementation of national policy is done in haste, without adequate backup, support and ownership by stakeholders, the chances that the result is a modified, educationally sound classroom curriculum may be limited. Successful realisation of a new curriculum is unlikely "unless the ground soil of public and political opinion has been very carefully tilled" (Atkins & Black, 2003, p. 26).

The discussion now examines in more depth the complexity and interconnectedness of the various components of curriculum exposed by the layers of meaning people assign to the term 'curriculum'.

2.3 Pedagogy and Curriculum

The International Association for the Evaluation of Educational Achievement (IEA) uses a framework for defining curriculum implementation that identifies teaching practice, or pedagogy, as a significant component of the overall implementation process (Orpwood, 2001). The framework consists of the:

- intended curriculum - that prescribed in official statements of the curriculum
- implemented curriculum – as actually taught or delivered in schools, and

- attained curriculum – as achieved by students.

McGee and Penlington (2001b) comment that recent focus in educational research on the close association between curriculum and teacher-student interaction recognises and acknowledges the contribution pedagogy makes to the curriculum actually experienced by students. Rather than national statements, curriculum is seen as a "lived, and evolving construction resulting from student-teacher interaction" (McGee & Penlington, 2001b, p. 1). Murphy (2003, p. 3) concurs, saying it is the implemented or "enacted curriculum" that determines the possibilities for learning rather than the specified curriculum of policy, and that it is teachers who organise this form of curriculum through their classroom practice. Their pedagogy influences the extent to which the intent of the specified curriculum is actually realised in the enacted curriculum. It follows that the closer the links between the theoretical basis of the intended curriculum and the pedagogical approach used by teachers the greater the likelihood the operational or student-experienced curriculum will match the intended. If, for example, a curriculum promotes the ability to do problem-solving in authentic situations as a curriculum aim, then students need learning experiences that allow them to develop this capability. Many authors suggest that pedagogical approaches based on social constructivist or sociocultural views of learning are most appropriate for achieving this sort of curriculum aim (e.g., Driver et al., 1996; Haigh, 2001; Harlen, 1999; Nuthall, 1997; White & Fredericksen, 1998).). According to social constructivist and sociocultural views of learning, appropriate teaching approaches would allow learners opportunities to construct the meaning and understandings needed within a social context and domain-specific science unit where conceptual change is shaped by prior knowledge (Duschl & Hamilton, 1998).

The IEA attempt to simplify the complexities of relationships inherent within the process of curriculum implementation by defining the intended, implemented and attained curricula can be problematic when it comes to interpretation (Orpwood, 2001). These IEA definitions of curricula appear to imply that teachers have control of the implemented curriculum, and students the attained or operational curriculum. Rather than deliberately choosing to learn or not learn, sociocultural views of learning (Murphy, 2003; Nuthall, 1997) would suggest that students influence the nature of the attained curriculum by the manner in which their learning dispositions, prior knowledge, and the social context give them access to the implemented curriculum. Eisner (1994) points out that the attained curriculum for students can, and often does, extend beyond the intended and implemented curricula to include elements of a 'hidden', or 'implicit' curriculum. To illustrate, students can acquire characteristics of schooling such as competitiveness, initiative or compliant behaviour that are not specified in the intended curriculum, nor intentionally taught by the teacher, because they are participants in "that persuasive and ubitiquitous set of expectations and rules that defines schooling as a cultural system" (p. 107). Erickson and Shultz (1992, p. 470) talk of the 'classroom underlife' (its informal social organization) as one of the fundamental processes occurring in classrooms that influence the educational experience of students, particularly relations between students and between teacher and students. These relationships have bearing on issues in classrooms to do with students' trust and feelings of legitimacy about their teacher's authority within the classroom, ownership of the taught curriculum, and their willingness to take risks in their learning.

2.3.1 The role of teacher in curriculum design and delivery

Not surprisingly, many authors see the role of teacher as pivotal to the learning of curriculum by students. The classroom curriculum students come to experience as reality is the responsibility and decision of the classroom teacher, and this, it is suggested, overshadows all other influences (Atkin & Black, 2003; McGee & Penlington, 2001b; Tytler, 2003). Of all the stakeholders with influence and interest in the nature of the student-experienced curriculum, teachers "are the only ones whose actions directly affect students' learning" (Harlen & Crick, 2003, p. 203). For example, research by Lederman (1999), in the classrooms of five high school biology teachers in the USA, found it was the teachers' instructional intentions that largely determined what occurred in classroom practice as the implemented curriculum. For teachers the design of this classroom curriculum is a highly complex task because teachers are subject to "multiple obligations and have to negotiate between a variety of internal and external determinants at all stages of the teaching process" (McGee & Penlington, 2001a, p. 10). In determining their instructional intentions, teachers therefore serve as intermediaries between the curriculum policy determined both outside and within the school, and the curriculum experienced by the students. However, because they possess their own personal belief systems, teachers also create their own layer of curriculum interpretation and pedagogical and assessment approaches, while carrying out their intermediary role within the constraints imposed by government policy and its agencies and the specific school context in which they work.

> Teachers don't merely deliver the curriculum. They develop, define it and reinterpret it too. It is what teachers think, what teachers believe and what teachers do at the level of the classroom that ultimately shapes the kind of learning that young people get. (Hargreaves & Fullan, 1992, p. i)

While individual teachers draw on a wide range of knowledge sources in their teaching decisions they often have their own strongly held conceptions and beliefs about what constitutes effective teaching and learning (Barnett & Hodson, 2001). They look to their own knowledge and belief base when creating the enacted classroom curriculum. In a study of 37 Taiwanese science teachers Tsai (2002) found strong alignment between the teachers' beliefs about science, and their views about teaching and learning. For example, if a teacher expressed an empiricist or logical positivist view of science then he/she was most likely to view teaching as the process of transferring knowledge and learning as a process of knowledge reproduction. In contrast a teacher viewing science as a human invention – a way of knowing about the natural world – tended to view teaching as helping students to construct knowledge and learning as students constructing personal understanding. Kang and Wallace (2004) in a study of the links between science teachers' epistemological beliefs, goals and practices found similar alignments, but teachers who allied relativist beliefs about science with knowledge having to be actively constructed by learners did not necessarily translate this into their classroom practice. These teachers often used transmissive methods because of the influence of contextual factors on their practice such as curriculum constraints.

Barnett and Hodson, (2001) also report on the influence of the context on teachers' practice and point out that each classroom is a unique educational environment, whose character is a function of a wide range of factors combining to produce each particular classroom culture. Differences between classrooms may arise from social, cultural or gender factors, or they may be related to teacher experience, age of students, or type of school, but factors contributing most to the uniqueness of each classroom relate to the students that comprise the class. The diverse circumstances and events of each classroom require the teacher to respond with ongoing, often complex, decision-making that impacts on the nature of learning (Leach & Scott, 2003). So teaching is not a simple, straightforward activity. McGee and Penlington (2001c) liken teaching to 'dilemma management' where teachers deal with a constant stream of problems by relying on their own personal knowledge, and coming up with the best 'solution' for the given circumstances.

2.3.2 The role of students in the experienced curriculum

Despite the intermediary actions of the teacher in 'repackaging' the intended curriculum Nuthall, (1997) suggests this adapted version is not necessarily the actual curriculum students come to experience and learn. Nuthall points to research that demonstrates how individual students mediate their learning on the basis of their prior knowledge, attitudes to learning, and involvement and ownership of classroom activities. These pre-existing ideas and dispositions influence students' interpretations and interactions with classroom processes, and appear to have considerable impact on the way students respond to the implemented curriculum, and what they consequently think and learn. Through language use and other social processes, Nuthall suggests students go through a form of enculturation as they develop their own sense of the cultural processes involved in learning. This understanding may facilitate or impede learning, because students use their personal awareness of how the classroom culture operates to reconcile their learning with influences and perceptions particularly impacting on them as individuals. The result is that students themselves have influence on what is taught and learned in classroom by creating their own interpretation of the teacher's intended or enacted curriculum. It is the total sum of all processes and activities going on in classrooms that needs to be considered if the nature of student thinking and learning is to be understood.

Carr et al. (2001) in a review of effective teaching and learning, note a wide range of factors reported to influence the affective aspects of learning in this mediating role of students such as self-esteem, acknowledgement for success and failure, feedback, motivation, and perceived successes and failures. Harlen and Crick (2003) single out motivation as the 'will to learn', encompassing many of the other affective factors in a complex mix of a learner's sense of self, engagement with learning and sense of control and efficacy, and willingness to exert effort to achieve a goal. To be motivated students needed to perceive good reasons for learning (Brophy, 1999) and appreciate the worth of what they were learning (Brookhart & Bronowizc, 2003). In a review of research into the impact of testing on students' motivation for learning, Harlen and Crick found that the motivation to learn can be "discouraged unwittingly by assessment and testing practices" (p. 204). Over-emphasis on testing can motivate students to strive for performance goals rather than learning goals, resulting in shallow rather than deep, life-long learning. Carr et al. (2001) stress the integral role of assessment in the achieved curriculum, and the capacity of various assessment purposes, such as formative and summative assessment, to influence the nature of student learning. This issue is considered in more detail in Section 2.4.

2.3.3 Teachers' response to new directions in curriculum policy

As well as their effect on classroom curriculum design, teachers' views and understanding of teaching and learning also have a very powerful influence on how they respond to new curriculum directions (Atkin & Black, 2003; Murphy, 2003; Powell & Anderson, 2002). If an introduced curriculum, and/or a new qualification structure, seeks to alter the content and approach to teaching and learning programmes in classrooms, teachers may well have to significantly change their classroom activities, values and thinking to achieve this (McGee & Penlington, 2001c; Murphy, 2003; Powell & Anderson, 2002). In this regard, teachers' opinions can 'make or break' a centralised attempt at curriculum change in terms of its intent.

Changing teaching practice, even when teachers are open to new ideas, is not a simple matter, since teachers face impediments inherent in the educational context in which they work (Loughran & Gunstone, 2003). In science education, for example, curricular aims have been continually evolving over the last 100 years in response to societal, economic and science needs, and this climate of continued change presents an ongoing challenge to teachers. Long-term cumulative effects of structural, social, psychological and normative factors on the everyday business of teaching, makes meaningful change problematic for individual teachers. The literature suggests that provision of detailed curricular text and guides on their own are not enough to effect teacher change if new curriculum policy represents major shifts in educational theory and content (Atkin & Black, 2003). For example, Murphy (2003) maintains that the failure of the 1980s British Design and Technology national curriculum to realise the potential for learning that advocates claimed for it, was in large part due to teachers' inability to accept the need to change pedagogical approaches to meet curriculum aims. Teachers cannot effect this kind of change alone: "To expect them to do this on their own without assistance is unrealistic and undesirable" (Bell & Gilbert, 1996, p. 44).

According to the literature, meaningful and long-term change of the delivered curriculum requires high quality professional development, conducted within learning communities where teachers are creative, active and reflective partners (Murphy, 2003). Effective professional development targets teachers' identified needs, and utilises concrete exemplars of classroom practice that they can identify with and implement in their classroom programmes in ways and at a pace appropriate to their levels of need and confidence (Barnett & Hodson, 2001; Loughran & Gunstone, 2003; Thornton, 2003).

2.4 Links between Assessment, Pedagogy and Curriculum

Once teachers can establish links between theory and practice, Atkins and Black (2003) claim that 'mature' teaching practice occurs when consideration turns to whether students are learning and how they are learning. These considerations form the basis of assessment for learning, and are crucial for informing teachers' work in designing and delivering the classroom curriculum to meet students' ongoing learning needs (Brookhart & Bronowicz, 2003).

2.4.1 Definitions of assessment from the literature

Broadly speaking, assessment can be defined as the process of gathering information about learners' work or performance and looking for evidence to use in decision-making (Peddie, 1992). These decisions may be about educational policy, curriculum and educational programmes, or individual students' learning (Gipps, 1994; Harlen & James, 1997; New Zealand Qualifications Authority [NZQA], 1996; Nitko, 1995; Sadler, 1989). Assessment impacts on many facets of teaching and learning, but the explanations and discussions that follow focus on how aspects of assessment relate to improving student achievement, since an important priority is that "a good assessment programme should encourage and assist learners" to achieve their learning goals (Peddie, 1992, p. 6).

2.4.1.1 Diagnostic, formative and summative assessment

The terms 'diagnostic', 'formative' and 'summative', when applied to assessment, refer to the function or purpose for collecting assessment evidence (Brookhart, 2001). Diagnostic assessment enables teachers to discover what students know, and can do, relative to the learning goals of the teaching unit (Gipps, 1994). Such information can allow teachers to identify specific difficulties or learning needs, discover their precise nature and scope, and plan learning activities to target the identified needs appropriately (MoE, 1994a).

Formative assessment is a term still under discussion in the literature (Black & Wiliam, 1998; Brookhart, 2001), and is yet to have a tightly defined and widely accepted meaning. Black and Wiliam (1998) in their literature review on this topic interpret formative assessment as "encompassing all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify teaching and learning activities in which they are engaged" (p. 7). Formative assessment then happens during teaching and learning, and the intention is to support learning via feedback, which 'feeds forward' into actions by teachers and learners to enhance learning (Bell & Cowie, 2001; Black & Wiliam, 1998). Effective formative assessment requires the teacher initially to clarify learning goals, task criteria and most importantly quality criteria by exemplification, and then through continuing dialogue with students as individuals or in groups usually by way of question-answer interactions (Torrance & Pryor, 2001; Clarke, 2001). Making certain task criteria explicit, like ways of working effectively together as a group, helps to make such practices more visible to students, something often assumed by teachers that students know inherently how to do. Allowing students time to improve on an initial attempt at a task provides opportunities for this "continuous clarification of criteria" (Torrance & Pryor, 2001, p. 624). Black (2003) comments that learning programmes must "provide opportunities to ensure that meaningful interventions that extend the pupils' understanding can take place" (p. 5), Inherent in Black's statement is the requirement that assessment information gathered in this fashion be used to inform teachers' future planning and actions. Thus, there is a dynamic element to this form of assessment (Glover & Thomas, 1999) because it is continuous and ongoing. Through teacher-student and student-student interactions (i.e., peer assessment) information is routinely gathered and communicated to students, telling them how successful something has been, or is being done, in relation to goals or reference levels depicting expected performance or achievement. Students also are informed about the actions they need to take to close the gap between actual and reference levels of achievement (Sadler, 1989). Such assessment serves to progressively shape and improve a student's competence to achieve a desired end – to scaffold learning (Torrance & Pryor, 2001). Thus formative assessment can be viewed as an integral part of the teaching and learning process (Clarke, 2001; Crooks, 2002; MoE, 1994a), and the review of research by Black and Wiliam (1998) reports that substantial learning gains can be achieved by students when teachers incorporate formative assessment into their pedagogies in classrooms.

Torrance and Pryor (2001) report on findings from the TASK Project (a research project into primary teachers' informal assessment practice based on classroom observation in the UK), which identifies two 'ideal-typical' approaches to formative assessment known as 'convergent assessment' and 'divergent assessment'. These two ideal types of assessment for learning represent the two ends of a spectrum of possibilities for teachers' practice, and seem to be associated with teachers' differing views on learning and the role of assessment in learning. Convergent assessment is concerned with the teacher finding out if a learner knows, understands or can do a predetermined thing, and the learning is measured from the point of view of the curriculum. This form of assessment is planned in detail, tends to use closed questions and tasks, and is more likely to be associated with behaviourist views of learning. On the other hand, for divergent assessment both student and teacher work interactively to discover what the learner knows, understands and can do, not from the point of view of the curriculum but from that of the learner. This form of assessment is not focused on the measurement of past or current achievement, but on the next learning steps for students. Divergent assessment is more closely aligned to constructivist views of learning and involves less planning and more open-ended questions and tasks. Torrance and Pryor (2001) note the use of one type of formative assessment by a teacher did not necessarily rule out use of the other in their everyday classroom practice.

The epitome of formative assessment in some writers' view (e.g., Black & Wiliam, 1998; Brookhart, 2001; Sadler, 1989) is the practice of self-monitoring. In contrast to feedback, where the teacher or a peer provides evaluative information, self-monitoring is a process where the learner generates such information independently of the teacher or peer. This self-generation of evaluative information, or internal feedback, is critical to the ongoing engagement of a learner with a learning task and the pattern of learning that evolves (Butler & Winne, 1995). Students only achieve this transition to self-assessment by gaining understanding of the learning goals, assessment criteria and actions needed to improve (Sadler, 1989). The capacity to self-assess provides a foundation for lifelong learning (Sadler, 1989) and is indicative of an effective learner (Black & Wiliam, 1998). Boud (1995) argued that self assessment was an important part of the curriculum at all levels, but Black and Wiliam (1998) reported that a focus on developing students' self assessment skills was not common teaching practice. McDonald and Boud (2003) found that few studies have been done in high school settings, so they undertook a study with students in their final year of study in high school to examine the effects of formal self-assessment training on student performance in external examinations. After intensive professional development in assessment, teachers integrated regular self-assessment training into their exam classes, and the exam results show that this training significantly improved the performance of students exposed to the training compared to their peers who had not received the training. McDonald and Boud feel this study provides strong indications that self-assessment can have a positive influence on students and their performance in examinations, but comment that further studies need to be done in terms of the impact of self assessment training on learning tasks other than those found in examination tasks.

Black (2003a) singles out peer assessment as an important complement to self assessment, and makes the comment that peer assessment has a number of advantages over feedback given by teachers. For example, students are more likely to actively seek a peer's assistance if they do not understand explanations given in class, often because the dialogue occurs in students' own language and there is less risk to their self esteem by not exposing their lack of understanding to a wider audience. Students tend to accept and take more notice of criticism of their work by peers. Also there is less personal risk in approaching a peer. By taking on these teaching and learning roles themselves students can come to see their own work more objectively, so assisting their self-assessment abilities.

Summative assessment is generally a more structured activity than formative assessment, and its principal goal is to do with gaining an overall view of the extent of learning that has occurred rather than helping ongoing learning (Brookhart, 2001). It takes place at certain intervals when a global indication of students' achievement has to be reported, and commonly relates to progression in learning against public criteria (Harlen & James, 1997). It may be concerned with summing up the achievement status of a student based on accumulating formative or smaller summative assessments, and has been equated by some writers with convergent formative assessment (Torrance & Pryor, 2001). It is more commonly seen as a means of checking up on progress against set criteria through tests or tasks at the end of a period of teaching and learning. Summative assessment is considered less potent at enhancing an individual's immediate or ongoing learning than formative assessment (Sadler, 1989), but it can still have positive effects if curriculum based and aligned closely to classroom instruction and the stated assessment criteria (Biggs, 1998). Since summative information is geared towards reporting at the end of a course of study (Peddie, 1992) it can influence decisions that have long-term ramifications for learners, such as the ability (or otherwise) to progress through levels of schooling, make subject choices or enter certain career paths. High stakes testing and examining for qualifications are instances where summative judgements are used to make certification decisions. These decisions often have important implications for students' future learning paths (Clarke, 2001; Sadler, 1989). There is growing awareness that testing for summative purposes can impact on students' motivation for learning, both positively and negatively. Practices like teaching to the test, and using class time to train students to pass tests via repeated practice have been shown to have negative effects on student motivation, while making grading criteria explicit and broadening the range of information used to assess the attainment of individual students can have positive effects on students' desire to learn (Harlen & Crick, 2003).

Where a teacher's classroom programme is based on particular curriculum such as a national curriculum, diagnostic, formative and summative assessments should be highly related to the learning goals of that curriculum (Nitko, 1995). It is important to appreciate that that these three types of assessment are not necessarily exclusive, indeed it may be difficult (and questionable) to attempt to differentiate between them

(MoE, 1994a). For example, diagnostic information can contribute to formative assessment in the initial stages of learning, and formative assessment can be used to diagnose learning needs during the learning process. Gipps (1994) describes diagnostic assessment as essentially a subset of formative assessment (p. 126), while in more recent research it often appears subsumed into formative assessment (Bell & Cowie, 2001; Brookhart, 2001; Wiliam, 2000). Summative assessment performed for one teaching unit can in turn serve some diagnostic and formative functions for ongoing learning in a subsequent teaching unit (Brookhart, 2001; Nitko, 1995).

2.4.1.2 Evaluation of assessment evidence

Evaluation of the extent of student learning involves making a judgement about students' assessment evidence by way of comparison (NZQA, 1996). There are three approaches to this comparison:

- self–referenced assessment (or ipsative referenced) where each learner's assessment evidence is judged against their own past performance.
- criterion or standards-based assessment where a learner's evidence is compared with a predetermined standard, often a fixed criterion or level of achievement (Peddie, 1992). In theory it is possible for all learners to reach the particular standard set, but in reality the number of students who actually achieve the standard(s) depends on the level the standard is set.
- norm-referenced assessment where each learner's evidence is compared with the achievement of others. There is an assumption that the abilities of students in the group being assessed can be sorted and ranked by plotting their results onto a bell-shaped curve or normal distribution (MoE, 1994a).

It is important to note that the New Zealand NCEA qualification utilises standards-based assessment information in judging learning achievements.

2.4.1.3 Validity, reliability, fairness and moderation of assessment evidence

To be truly trustworthy and useful, assessment information needs to be gained through processes that have high degrees of validity, reliability and fairness (MoE, 1994a). Validity in assessment is to do with fitness for purpose and usefulness. Consequently the appropriateness of an assessment task for assessing particular learning outcomes, and the degree to which it measures what it is supposed to measure are important considerations if validity is sought (Gipps, 1994; MoE, 1994a; Peddie, 1992; Orpwood, 2001). So too is ensuring that the inferences users make of the information produced are justified (Black, 1995). This "consequential validity" (Gipps, 1994, p. 61) is the essence of formative assessment, because such assessment "cannot claim to be formative until it demonstrably leads to action for improved learning" (Harlen & James, 1997, p. 371).

Reliability is concerned with the consistency with which an assessment task or method measures what it is designed to measure over different activities, markers and time (Gipps, 1994; MoE, 1994a; Peddie, 1992). Consistency (or replicability) of assessment results implies that the information obtained can be relied on to give an undeviating and accurate picture of what was measured (Cohen et al., 2000). While formative assessment is private and focused on the needs of the learner, summative assessment tends to be public in response to external requirements such as accountability or accreditation or for national monitoring (Brookhart, 2001). Summative assessment therefore requires assessment methods that are as reliable as possible (Harlen & James, 1997; Orpwood, 2001), allowing assessors to make the same judgement on different occasions in relation to similar skills or knowledge criteria. Formative assessment on the other hand need not be overly concerned with reliability since the assessment information is used to inform teaching and help learning in the specific context in which it is gathered (Black, 2001; Harlen & James, 1997). The usefulness of formative assessment lies in informing teaching, and helping students learn in a way that enables them to apply the same skills and knowledge to different contexts. The information gained from formative assessment and used to enhance learning shows greater reliability if students can then successfully demonstrate their understanding in different learning contexts.

To be reliable assessment methods need a high degree of validity, but validity by itself does not guarantee reliability (Peddie, 1992). Ensuring reliability can necessitate assessing more than once, to take into account all the chance factors that may affect the results at the time of each assessment. Sometimes reliability may have to be sacrificed for validity. For example, a multichoice test may offer more reliability because it is more objectively scored, but an essay may assess the required knowledge and skills more validly if students' ability to synthesise, critique or think creatively is being evaluated.

Fairness in assessment is the requirement to eliminate bias and to promote equity and inclusiveness in assessment methods so no individual, or cohort of students, are advantaged or disadvantaged over others (Gipps, 1994). Keeping assessments fair may, however, compromise validity and reliability. For example, using multiple-choice questions, for example, to reduce subjectivity and improve validity in assessment can lead to concerns about gender fairness (Lokan et al., 1999). Many studies show that males perform better when multiple-choice components are added in areas where gender performances have been previously deemed equivalent. Conversely, females perform better in essay-type questions. Broadening assessment programmes to include a wider variety of tasks may thus be a means of addressing issues of validity, reliability and fairness.

Assessment moderation is a process that can be undertaken by teachers to maximise the validity, reliability and fairness of assessment practices in order to achieve the consistency and quality of assessment practice (Gipps, 1994). It is commonly encountered in assessment that leads to certification and qualifications, and involves checking for comparability at different stages of the assessment process, such as task construction and marking. Assuring that an assessment is roughly the same as comparable assessments by other teachers, or at other institutions, is an example of moderation (Peddie, 1992). National curriculum statements and prescriptions are essential elements of moderation processes in that they define what is to be commonly learned by all students (MoE, 1994a). Checking for validity and consistency can take place informally between teachers in discussion, or formally through common assessment tasks, standard assessment tasks, common grading systems, and shared or group marking. Moderation is especially significant if the achievements of students from unlike groups need to be compared, for example, students with different teachers or from different schools or regions.

Validity, fairness and reliability of assessment information requires high levels of professional competence on the part of teachers resulting from careful training and prolonged experience (Atkins & Black, 2003). Manageability is concerned with ensuring that any assessment in the classroom does not place undue stress on the teaching and learning processes. Tensions commonly arise as professionals seek to address concerns of validity, reliability, fairness and manageability in assessment tasks (MoE, 1994a), and final decisions inevitably involve compromise to some degree.

2.4.2 Tension between formative and summative assessment within qualifications

International research suggests that when formative and summative assessment are both included in national assessment policies, the way in which they are related to one another causes a blurring of the distinction between their purposes (Harlen & James, 1997). While the terms are initially distinguished in terms of purpose and timing, subsequent statements can suggest that assessments originally made for one purpose can be combined to meet a different purpose. The New Zealand Curriculum Framework (MoE, 1993b) appears to reflect this tendency:

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

The statement infers that information originally collected for formative purposes is to be accumulated and used also for summative purposes. Teachers will be required to help their students on the one hand and make judgements about the quality of their work on the other (Atkins & Black, 2003). This dual role is not easy for teachers. For example, summative functions frequently call for norm-referenced scores, so a range of complex formative assessment information must be combined into a single score so that individuals' achievement can be ranked (Black, 2003a). Further tension can arise if teachers come to perceive that one form of assessment has greater importance or status than another (Black, 2003b; Carr et al., 2001; Harlen & James, 1997). When high stakes assessment, such as national testing or qualification tasks, is also involved this blending of purpose can be to the detriment of formative assessment, as the demands of summative assessment tend to dominate (Harlen, 1998; Preece & Skinner, 1999).

Research suggests that in high stakes assessment situations many teachers are inclined to focus their pedagogy on 'teaching to the test' using continuous summative assessment in the guise of formative assessment (Atkins & Black, 2003; Carr et al., 2001; Harlen & James, 1997; Harlen & Crick, 2003; Reay & Wiliam, 1999), so the potential for students to learn from genuine formative assessment is hampered. An example is the introduction of national testing in England, Wales and Northern Ireland for pupils at ages 7, 11 and 14. Since the results were used to compare schools' performances in league tables, the consequences were that teachers tended to give more of their attention to what was tested. Not only was the formative function of assessment sidelined, but also because the tests focused on a very limited measure of performance, knowledge and understanding, less attention was given to practical work and investigatory skills in subjects like science (Harlen, 1998). In Scotland, where national testing does not occur in primary science, more emphasis is given to formative assessment in supporting documents for teachers, and teachers have the latitude to explore ways of incorporating the practice into their teaching. In contrast, it is interesting to note that in a study in the USA Vogler (2002) reports teachers' practice of 'teaching to the test' having a positive spin-off for student learning. It seems that if high stakes performance assessments use techniques that promote broader and higher levels of thinking, such as open-response questions, creative/critical thinking and inquiry/ investigation, then teachers' instructional methods change to include more that are regarded 'best practices' by educational researchers. Teachers' instructional practices begin to include more child-centred, experiential, authentic, and reflective approaches.

However, the effects of high stakes assessment on learning practices and the tensions that can be set up when assessments need to serve both formative and summative purposes cannot be ignored (Black, 2003b; Black & Wiliam, 1998; Brookhart, 2001). The teacher, acting as both support and judge of student's learning could place the effectiveness of formative assessment in jeopardy if students perceive their every move as being appraised summatively (Bell & Cowie, 2001). Students quickly pick up from their teachers what is valued, and the criteria the teacher is using to judge their work for summative purposes (Harlen & Crick, 2003; Reay & Wiliam, 1999). Where that focus is performance-oriented, students' attention will inevitably turn away from learning processes to concentrate on giving responses that will 'pass the test', rather than developing the deeper skills of application, analysis and synthesis.

Conversely, teachers may feel under undue pressure to secure good results for their students, and hence their schools, which could threaten the reliability of their assessments (Black, 2003b). Teacher inexperience and/or lack of expertise can cause sharp division between normal classroom assessment work, and summative exercises can affect the validity and reliability of their assessment practice (Yung, 2000, cited in Black, 2003b).

While there is considerable evidence that suggests formative assessment does not sit comfortably within qualification structures, there is a model in existence that has been reported to achieve this for nearly 30 years (Black, 2003a). The Queensland Senior Certificate Examination System focuses on the formative function of assessment, and uses teachers' judgements of students' performance against specific criteria as a basis for accreditation. Students receive feedback and feedforward information from continuous formative and summative assessments over a two-year period. An external governmental system, comprising review panels made up largely of teachers, has the task of maintaining comparability of curriculum and standards in schools and classrooms across the state. This overseeing by the state, in Black's view, protects the credibility of the qualification without jeopardising the primary formative function. However, Black comments this alignment of formative and summative functions was only achieved through considerable reorganisation of the schooling and qualification system, and changes in approach and understanding of the issues by schools and teachers.

An important aspect contributing to the validity of summative assessments is that the meaning and significance of the assessment outcomes are communally shared, such that "the same score must be interpreted in similar ways for different individuals" (Wiliam, 2000, p. 11). Arriving at a summative judgement that assures some common standard has been met, requires reviewing of assessment evidence against known criteria (Harlen, 1999). For this purpose the distinction must be made between making a summative judgement using judgements already made (continuous-summative), and that made against criteria using the accumulated evidence taken as a whole. The summative judgement should not be an averaging of levels previously scored, but rather a report on learning progress to date. This involves the separating of the elicitation of assessment evidence from its interpretation (Wiliam, 2000), and interpreting the evidence differently for different purposes. The teacher goes back to information originally collected and used for formative purposes, and interprets that primary evidence retrospectively, this time for summative purposes. Consistent judgements can be aided through the processes of moderation and exemplification.

There is now a growing belief in educational circles that ways need to be found to make assessment for learning compatible with assessment for qualifications (Black & Wiliam, 1998; Black, 2001, 2003a; Brookhart, 2001; Harlen & James, 1997; Wiliam, Lee, Harrison & Black, 2004). Excluding formative assessment from qualification would be a backward step by denying a powerful learning tool to students. Teachers' best efforts can be frustrated and students' self esteem and motivation to learn threatened if all assessment is regarded as summative (Black, 2003a). If student achievement (in terms of higher levels of learning and skills over a broader range) is the goal of qualifications, then the proven potential of formative assessment to improve student learning should not be ignored (Black & Wiliam, 1998; NZQA, 2001).

As Wiliam (2004) points out:

> Whatever a logical analysis of the problem suggests, rather than adopting entrenched positions on one side or other of the debate, we must refuse to accept the incomparability of 'summative' and 'formative' assessment. Instead we must find ways of mitigating that tension, by whatever means we can. (p. 4)

The review now turns to the international science education scene to gauge how the interplay of current thinking and practice in the key areas of curriculum, pedagogy and assessment is impacting on trends in the teaching and learning of scientific inquiry in national science curricula.

2.5 International Trends in National Science Curricula

Neyland (1995), in his account of influences on mathematics curricula in the 1990s commented that two opposing world views, or set of beliefs, were influencing the development of curricula in western countries. The 'technocratic view' of the world sees the gaining of knowledge as individualistic and empirical, requiring a reductionist or atomistic approach oriented towards prediction and control. In contrast the 'holistic', 'organismic' and 'systemic' perspectives see knowledge requiring active construction from both observable and non-observable elements, and dwelling in society as well as in individuals. The continuing influence of these two perspectives is reflected in recent literature reviews of trends in science education (Bell, 2005; Carr et al. 2001; Hipkins et al. 2001). The review by Hipkins et al. in particular provides a more specific summary of the predominant learning theories and research trends in science education that have influenced national curricula development over the last 40 years. In their view the key learning theories holding sway today are: the behaviourist views which dominated in the 1960s and 1970s; the developmental theories of the 1970s; personal constructivist theories from the 1980s; and the social constructivist and socio-cultural views of learning that emerged during the 1990s. These theories have spawned a variety of teaching and learning approaches, including discovery and process approaches to learning, teaching for conceptual development and metacognitive thinking, negotiated curriculum and setting science in authentic contexts.

In terms of emerging international trends in science education, Carr et al. (2001) report that most countries now have detailed and mandated curricula with many having some form of national standards describing concepts for students to learn. These standards are often expressed in the form of behavioural outcomes (Orpwood, 2001). Carr et al. (2001) note the influence these standards have on teachers and students differs from country to country and is not well researched. However, they observe that no one country is happy with existing programmes of science education, even those who score highly in international comparisons of educational achievement. Science educators dissatisfied with aspects of current programmes, argue that school science does not reflect real science as currently practiced, and students are not gaining a true understanding or appreciation of the nature of science and how it works (Gott, Duggan & Johnson, 1999; Hurd, 1997; Lederman, 1999; Mayer & Kumano, 1999; Millar & Osborne, 1998; Osborne, 2002). Duggan and Gott (2002) cite over-emphasis on content, students' perceptions of science as difficult and irrelevant, and the public's general lack of scientific literacy as evidence of the inappropriateness of current science curricula. To address some of these concerns, Carr et al. (2001) report that all countries appear to be seeking reform of their curriculum goals to more accurately portray authentic scientific inquiry and encourage creativity, problem-solving, and student enthusiasm for learning.

2.5.1 The nature of authentic scientific inquiry

Scientific endeavour is not about getting ‘right answers’, but rather a social practice that involves participants in interpreting, negotiating and justifying in order to build believable and plausible explanations about how the physical world works (Haigh & France, 2005; Wallace & Louden, 2002). The scientific knowledge that results from this social activity are potentially different from other forms of human knowledge and ways of knowing (Sandoval, 2005) in that: scientific knowledge is constructed with a dialectical relationship between theory and observation; there are different forms of scientific knowledge such as theories, laws, hypotheses and models; scientific methods are diverse; and scientific knowledge varies in its degree of certainty. Thus “science in ‘real’ laboratories is conducted within a social milieu of interpretation, justification and argumentation” (Wallace & Louden, 2002, p. 37). Within these communities of practice scientists operate from positions that are linked to particular paradigms. These paradigms provide the sets of beliefs, tenets and premises that underpin the social behaviours of scientists, including the process skills they employ.

In a survey of those mental and physical skills accorded the title of ‘scientific process skills’, Harlen (1999) found agreement in the literature that these skills were; “in one form or another, abilities related to identifying investigatable questions, designing investigations, obtaining evidence, interpreting evidence in terms of the question addressed in the inquiry and communicating the investigation process" (p. 129). While these process skills could be said to be generic to all investigative work, Atkin and Black (2003) maintain they are only termed 'scientific' if applied in the context of science and informed and guided by scientific theory: "The processes only become scientific when they utilize scientifically significant content and embody scientific purpose" (Hodson, 1992, p. 122). This requirement to give meaning to scientific inquiry by linking science concepts with the purpose of an investigation is a key means of differentiating scientific inquiry from other forms of investigation (English & Wood, 1997). Scientific inquiry is the process by which new understandings and knowledge in science are generated and validated.

In real science inquiry scientists are frequently presented with open-ended problems which Reid and Yang (2002) define as problems where there is no data, known method or established goal. In such situations all three components have to be developed by scientists in order to tackle the problem. Reid and Yang point out that successful open-ended problem solving depends on the knowledge and experience held by the people involved, and their ability to draw on appropriate and relevant information. This observation recognises both the theory and experiential-driven nature of scientific inquiry. Hodson (1992) contends that scientists do this intuitively, using their own personal theoretical constructs and tacit knowledge of how to do science. He describes how this 'art and craft', or 'connoisseurship', gives scientists the "capacity to use theoretical and procedural knowledge in a purposeful way to achieve certain goals" (p. 133), and believes this only comes with experience.

2.5.2 The case for inclusion of authentic student investigations in science curricula

The 'doing of science' by students is gaining new prominence in science education worldwide (Atkins & Black, 2003; Keeves, 1998; Sandoval, 2005) as a means of achieving the goals of curriculum reforms currently being implemented. This 'doing of science' is intended to involve students in experiencing the procedural and conceptual knowledge required to carry out scientific inquiry in a manner that mirrors the authentic practice of scientific communities (Duggan & Gott, 1995; English & Wood, 1997; Haigh & Hubbard, 1997; Hodson, 1995). Supporters of this authentic approach believe that much of what a scientist does is based on his/her tacit or intuitive knowledge, which is expertise gained cumulatively through the experience of 'doing science' in holistic investigations in many different contexts. To develop knowledge and expertise along a similar line to scientists, many international science educators take the position that students can achieve this best through participation in independent, genuine investigations where the solution to the problem is not obvious (Atkins & Black, 2003; Chin & Kayalivizhi, 2002; Hodson, 1992; Jenkins, 1996; Luft, 1999; Orpwood, 2001; Powell & Anderson, 2002; Weinburgh, 2003). Curriculum designers in most countries are approaching this new emphasis on authentic inquiry in the context of practical work (Carr et al., 2001), although some recognise great potential for problem-solving by other medium such as computer simulations (White & Fredericksen, 1998), or through language, literacy and argument (Osborne, 2002).

Science educators who argue for inclusion of student investigations in science programmes do so on the basis of the following educative reasons. Investigations are believed to:

- Motivate students' interest and desire to learn science. By carrying out investigations into phenomena and problems of personal interest and pertinence, students are more likely to own and direct the activity, and so gain experience of what genuine scientific inquiry can entail (Hipkins & Booker, 2002; Hughes, 2004; Jenkins, 1996). Deboer (2002) cites genuine and relevant inquiry tasks as valuable pedagogical tools because they can provide powerful intrinsic motivation for students to learn. He maintains investigations that are related to the science content being taught encourage student engagement and provide vital learning links. Also many of the attitudes and dispositions associated with genuine scientific inquiry, such as curiosity and personal search for meaning, are also those of autonomous and self-motivated learners. Reid and Yang (2002) believe the confidence needed to take the cognitive risks that open investigations often require can be developed through experience, especially successful experience.
- Enhance students' conceptual and procedural understanding of science. Broadly speaking, these two aspects encompass the learning of science and how science works (English & Wood, 1997; Garnett & Garnett, 1995; Roberts, 2004; Skamp, 2004). Duggan and Gott (2002) describe conceptual understanding as a knowledge base of the big ideas in science (such as the laws of motion, evolution and chemical change), and the scientific facts that underpin these ideas. Procedural knowledge is the 'thinking behind the doing' of science - the ability to put together a solution to a problem rather than to

follow a recipe. To think this way Duggan and Gott (2002) explain that scientists possess abilities known as ‘skills’ and ‘concepts of evidence’. Skills are the mechanical aspects of doing science like using measuring instruments or creating tables and graphs, while concepts of evidence are ideas concerned with how best to collect valid and reliable evidence in order to convince others of the implications of the data. Duggan and Gott believe it is students’ ability to acquire these concepts of evidence, especially an appreciation that evidence must be able to withstand public scrutiny, that leads to them becoming scientifically literate (i.e., obtain an awareness of how science produces and validates new knowledge).

- Apprentice students into the culture of science. Student investigations are seen as an effective way to gain first hand experience of science’s system of thinking and working to create new knowledge (Atkins & Black, 2003; Collins, 2004; Duschl & Hamilton, 1998; Hart et al., 2000; Hipkins et al., 2002; Jenkins, 1996; Luft, 1999; Powell & Anderson, 2002; Weinburgh, 2003)
- Enable students to develop an understanding and appreciation of the nature of science. The authentic first hand experience that students gain in investigations can facilitate meaningful learning about scientific ways of knowing and the relationship of science with technology and society (Driver et al., 1996; Duggan & Gott, 2002; Hipkins & Booker, 2002; Orpwood, 2001).
- Improve students’ thinking and learning capabilities. Investigations can serve as opportunities to encourage collaboration, and foster creativity, higher order critical thinking and problem solving (Chin & Kayalivizhi, 2002; Duggan & Gott, 2002; Haigh, 2003; Reid & Yang, 2002; White & Fredericksen, 1998).

2.5.3 *Pedagogy for authentic scientific investigation*

Harlen (1999) suggests the main strategies that teachers can use to help the learning of science process skills for inquiry work include: providing students with an opportunity for using process skills; encouraging critical review by students of their own performance and identification of ways in which they can improve that performance; teacher feedback that focuses on the quality of the work, not on the person; providing students with exemplars which meet the criteria of quality; engaging in metacognitive discussion about procedures to facilitate transference of skills to other contexts; and teaching the techniques and language needed as skills advance.

Student participation in genuine scientific investigations gives them the opportunity to develop and use process skills and ideally should involve them in a form of problem-solving where the solution to the problem is not obvious (Duggan & Scott, 1995). Authentic inquiry would require students to draw on their existing science ideas to analyse a problem, plan a course of action, carry out the plan to obtain information that they can analyse, interpret and evaluate to reach a conclusion (Garnett & Garnett, 1995). Finally their inquiry and findings need to be communicated in some form. Just as scientists build scientific knowledge, Hipkins et al. (2001) report that the utilising of scientific ideas for scientific purposes is widely believed to play a crucial role in the development of science understanding for students. Many authors assert science process skills cannot be separated from the body of science knowledge that underpins the learning and application of science (e.g., English & Wood, 1997; Harlen, 1998; Hodson, 1992; Leach & Scott, 2003; Luft, 1999) – they are integral rather than discrete components of the curriculum. The linkage of content with process skills therefore needs careful consideration and attention in curriculum and teaching (Atkins & Black, 2003; Harlen, 1998; Keeves, 1998). Proponents of this view call for explicit recognition in science curricula of the conceptual frameworks within which scientists operate, and the manner in which these frameworks influence how scientists come to understand and generate new knowledge and understanding. Such a curriculum would provide a model for student learning in science that mirrors authentic scientific knowledge building. This model, in practice, would see students making explicit the theoretical background or perspective that provides their conceptual framework or rationale for planning, performing and evaluating the findings of their investigation (Hodson, 1996; Lake, 2004).

If teachers want to facilitate students' participation in authentic science, to give them insights into aspects of science practice and how scientific knowledge is established, then Hart et al. (2000) feel teachers need to be very focused and explicit about their pedagogical intent with students. Students need to be made aware of the 'purposes' for engaging in investigations so they have clear learning goals. To make their pedagogical intent explicit to students, teachers need to be fully cognizant of the content they are teaching. Gott et al. (1999) suggest that most teachers are unaware of the significance of procedural understanding in science (how scientists think and work), and consequently it is not explicitly taught to students. They argue that this particular content needs to be defined and known by teachers so it can be targeted in their teaching, for example, providing activities and exercises to develop understanding of the concept of repeatability.

Drawing on her research findings into student investigations, Haigh (2001) describes a pedagogical model for student investigations that views the learning process as one of enculturation. In this model the teacher plays the role of expert in the community of practice known as ‘science’, in which students are the novices or apprentices. For this model to succeed the teacher will need to be highly skilled and knowledgeable about science since the teacher’s role is vital in supporting students as they acquire the tools and concepts needed to carry out scientific investigations in authentic situations. Haigh (2001) recommends pedagogical techniques such as role-modelling, analysing of structured experimental methods, questioning, and providing feedback and feedforward for facilitating this process of enculturation of novice students into the ways of science.

2.5.4 The reality of current pedagogy and student learning

In reality classroom pedagogy that facilitates links between authentic scientific practice and that experienced by students appears to be scarce. Commentators on current classroom practice note practical work in school science often bears little resemblance to inquiry in real science, especially where represented as an experience of ‘the scientific experience’ by teachers (Driver et al., 1996; Hipkins & Booker, 2002; Hodson, 1990, 1996; Woolnough, 2001). Many authors acknowledge the introduction of student investigations into science practical work, but observe a continued focus on conceptual learning and reliance on recipe-style laboratory work in most classrooms, especially in secondary schools (Chin & Kayalivizhi, 2002; Garnett & Garnett, 1995; Hipkins et al., 2002; Hodson, 1996; Haigh, 2005). In these practical science sessions little attention is given to problem-solving, design and critical evaluation. Gott et al. (1999) report "that most pupils can carry out practical tasks adequately but that few can understand, interpret or evaluate their data" (p. 100). Gott et al. believed that because these aspects of procedural understanding underpinning the doing of science were not directly taught to students, only the very able students were able to pick up these concepts.

Hodson (1996) believes that classroom practice is being influenced more by approaches to teaching and learning, such as discovery learning, the process approach and constructivism, than those promoting scientific activity as practiced by science. As a consequence he believes the nature of scientific inquiry continues to be misrepresented to students. Hipkins and Booker (2002) agree, claiming teachers mask the essential features of science that make it science by focusing on a particular pedagogical approach – for example, adopting a discovery approach denies students the opportunity to interpret their observations from the basis of scientific theory and concepts as scientists do. Hipkins and Booker comment that there can be no one best way to teach scientific inquiry because scientists actually work in a variety of ways to achieve their ends. Other writers agree (e.g., Hughes, 2004; Jenkins 1996; Roberts, 2004; Watson et al., 1999), and point to the methodological differences between the sciences, calling for recognition of the wide variety of investigative modes that practicing scientists actually utilise in real life problem solving. Mayer and Kumano (1999) believe students' investigative experiences should be broadened from fair testing to the sort of process approaches used in systems-based sciences such as ecology, geology, astronomy and meteorology. Watson et al. (1999) identify five other kinds of investigation commonly occurring in science, including: classifying and identifying, pattern seeking; investigating models; exploring; and, making things or developing systems.

The following paragraphs cite examples of international research about classroom practice that support these conclusions about the manner in which investigative science is currently being taught in schools world-wide, and the nature of the student learning. These examples also identify some specific pedagogical and learning issues that arise during investigative work, and include some instances of pedagogical strategies that are being developed to address these issues and promote more authentic student investigations in science.

In a large-scale study investigating the quality of science teaching and learning in Australian schools findings indicate that traditional closed laboratory exercises, where the solution to the problem being investigated is known, are still the norm in Australian secondary schools (Rennie et al., 2001). Similarly, Atkins and Black (2003) observed in their studies of classrooms in the UK and USA that only a small proportion of teachers taught in ways that enable students to engage in investigations reflective of real science. Driver et al. (1996) agree, adding that rarely is science portrayed as a social enterprise. It is not common practice, for example, to use teaching strategies that enable groups of students to evaluate a scientific theory or model in their investigations. In the UK, Tytler and Swatton (1992) suggested a factor limiting authentic scientific activity for students was the ‘control of variables’ (or fair testing) model of science investigation, embodied in Attainment Target 1 of the National Curriculum. In their view the tightly controlled structure of the target, focused on fair testing, was distorting the work of practicing scientists. They also feared the fair testing approach would lead to a model of investigative science that was assessment driven, counterproductive to creativity and potentially demotivating for students. Some four years later, despite subsequent refinements to the investigative target to make it less specific, this concern appeared justified when Laws (1996) reported that the classroom reality of practical inquiry remained 'disappointing'. He acknowledged the potential of inquiry based learning for giving students insight into the nature of science, but warned student investigations that were truly comparable to the complexity of real scientists' activity were going to be difficult to achieve. Such concerns are reported elsewhere with White and Frederickson (1998) finding little evidence of genuine science investigations in American classrooms. White and Frederickson comment that many of their teachers lack expertise in inquiry-based research simply because teachers themselves have never been practicing members of research communities. It is therefore difficult for such teachers to apprentice students into a research culture that resembles that of authentic science.

Even when the pedagogy actively seeks to promote authentic inquiry students' learning outcomes may not reflect the teaching goals. In a study of laboratory work and its impact on student learning in two Australian schools, Berry et al. (1999) found that in open investigations students' learning was fairly superficial, and focused on following a procedure and finishing the task. Factors that did cause students to become more mentally engaged in their tasks included: having sufficient relevant content knowledge, ownership through input into the design or implementation of the task; time to select and/or develop their own laboratory work; and awareness of the point of their investigations and the purpose of the investigations in their science learning. In subsequent research Hart et al. (2000) found additional evidence of the benefit of making the pedagogical purpose clear to students, where findings from their classroom study demonstrated that this knowledge helped students derive understanding of how scientists work during their practical investigations.

Research identifies scaffolding of student learning as essential in promoting authentic scientific investigative work. For example, Haigh and France (2005) report from research in New Zealand into student investigations in biology that students experience difficulty with open-ended investigations, unless the learning has been carefully scaffolded in the transition from closed to open inquiry. Loss of confidence by students, as they come to realise the complexity of authentic inquiry, and teachers' insecurities can work against positive outcomes from investigations unless due attention is paid to these aspects. When carefully structured support was provided in this study (Haigh & France, 2005), the biology students achieved a wider range of scientific learning outcomes than they appeared to gain in more traditional practical work. Hipkins and Booker (2002) describe an attempt by a NZ science teacher to enable her students to experience authentic investigative work for Science Achievement Standard 1.1 *Carrying out a practical investigation with direction* (which will be revisited in greater detail later in this chapter), but they do not comment on the nature of the learning outcomes for these students

Writing from the USA, White and Fredericksen (1998) recognise this need to scaffold learning in scientific investigation, and have developed a metacognitively-focused, inquiry-oriented curriculum that relies not just on real-world materials but also on computer models. In inquiry and reflective cycles that closely mirror the way scientists work, students' performance and achievement in science have improved significantly. White and Fredericksen believe the software modelling tools allows a wider ability range of students to work with and construct conceptual models than with just practical work alone. This process of scaffolded inquiry, reflection and generalization promotes the development of the necessary metacognitive knowledge and skills. As far back as the early 1990s Hodson (1991) strongly advocated computer-based learning, specifically computer simulations, to give students more authentic opportunities to acquire the thinking skills of creative experimental scientists. The Creativity in Science and Technology (CREST) programme used in some New Zealand schools (Coles & Coles, 1997; Hipkins et al., 2002; Hume, 1995) also encompasses a reflective cycle that encourages students to engage in metacognitive processes. Students reflect on their thinking processes as they perform open-ended investigations into topics of their own choice, and with practicing scientists as mentors. In an evaluation of the CREST programme Davies and France (2001) report that participating students developed investigative abilities and scientific knowledge well beyond that achieved in school programmes.

Once students begin engaging in authentic science inquiry as part of schools' teaching and learning programmes, then the question of assessment inevitably arises. New pedagogical and learning approaches usually need corresponding modifications to assessment practices.

2.5.5 Assessment for authentic science investigation

To align assessment practice with curriculum goals and carry out valid and useful assessment, Hodson (1992) argues activities must be based on a true model of the nature of science. Assessing scientific process skills as isolated, decontextualised, individual skills would raise validity issues in his view. This would, for example, reinforce the image of scientific inquiry as a linear, logical and unproblematic process, which Hodson maintains is not an accurate portrayal. In his opinion, and others (e.g., Black, 2001; Haigh, 2001), students should therefore be assessed, formatively and summatively, in the context of whole investigations so they can learn and demonstrate the knowing of what, when, why, where, and how to investigate; the gaining of laboratory skills; and the performing of the investigation. To enable holistic assessment of process skills the approach will need to be logistically school-based with the teachers having a key role.

Assessment of scientific process skills is not easy to perform in the reality of a classroom (Harlen, 1999; Keeves, 1998; Mathews & McKenna, 2005), especially when teachers focus on the development of scientific concepts and knowledge. To make high quality, holistic judgements teachers themselves have to possess deep understanding and appreciation of the nature of scientific investigation. The effectiveness of the teacher's formative and summative assessment will be compromised if he/she is not knowledgeable of the broad principles and key ideas they are aiming for in the learning of investigatory science (Harlen, 1998). Research indicates that many teachers have a limited and rather rigid view of how science is practiced, often associated with the so-called 'scientific method' (Weinburgh, 2003). This superficial understanding could reflect the teachers' general lack of personal experience of conducting research in real world environments, and their reliance on textbooks that promote 'the scientific method' for much of their science content knowledge.

Not surprisingly, in light of these findings, the development and use of assessment tools to validly monitor student learning in inquiry have been almost universally slow. Orpwood (2001) suggests the reasons may include:

- the lack of assessment technology to assess students' abilities to conduct scientific investigations
- scepticism amongst some in the education research community about the reliability of such tools once developed
- lack of credibility in the form of assessment by the public, and
- an unwillingness amongst teachers to change from traditional assessment methods.

Gott and Duggan (2002), in an overview of problems associated with the summative assessment of practical work in science in the UK see no easy solution to the problem of assessing scientific inquiry validly and reliably. UK experience with national testing in the late 1980s and early 1990s demonstrated "that any attempt to implement radically new assessment procedures is bound to fail if it is imposed without allowing teachers time to understand, and to influence, the changes" (Atkins & Black, 2003, p. 132). It was in the 1980s that performance assessments first appeared in UK schools, but the 1990s before they began to be widely accepted as appropriate measures of the new curriculum goals related to inquiry. Orpwood (2001) reports, however, that by the end of the 1990s little was available by way of guidance and exemplary materials to assist teachers in performance assessment. The growing significance placed on national assessment in the UK, particularly its accountability functions, and the continued focus on traditional standardised paper-and-pencil assessments, put teachers convinced of the value of inquiry-based learning under pressure. Orpwood suggests this pressure is proving difficult for teachers to resist, and learning based on authentic investigations is suffering as a result. Deboer (2002) contends there is potential for tension to occur when student-centred inquiry is attempted within a standards-based environment, because the freedom of students to inquire is curtailed by the need for the teacher (and student) to cover the subject matter of the standard. Student learning is therefore directed at the subject matter of the standard. The more specified the content the more potential for tension. However, the more general the learning goal of a standard, the more difficult teacher accountability and assurance of quality learning outcomes become. Deboer proposes a compromise where important concepts fundamental to the understanding of science are identified and stated as broad expectations. This gives teachers and students latitude to choose content and pedagogical approaches appropriate to students' inquiry interests and learning needs

In a detailed study of practical work carried out by 14-16 year old students working within the GCSE assessment scheme Keiler and Woolnough (2002) report the focus by teachers and students was on performance rather than learning. Students excelled by playing the 'rules of the game' to meet assessment requirements, "including doing extraneous, possibly erroneous or fallacious work" (p. 87) to earn marks. In recent work Roberts and Gott (2004) investigated the effect that using templates for planning and reporting investigations in the standards-based UK assessment system had on learning outcomes for students. These templates, which act like shells or blueprints for assessment developers (Solano-Flores, Jovanovic, Shavelson, & Bachman 1999), were introduced to improve reliability across a range of subject matter and contexts in the assessment tasks, and the exemplification of the assessment criteria in particular contexts has led to a repertoire of standard items building up. Roberts and Gott describe the exemplars, which serve important moderation functions, as almost a 'seen exam' situation. They quote the situation for Science Attainment Target 1 of the National Curriculum in Britain where "assessment has become routine, with a limit on the number of cases assessed. In some instances, Sc1 [Science Attainment Target 1] coursework has become so formulaic that performance is more akin to the recall of a complex protocol than the creative solution of a problem." (p. 104).

In investigative work new assessment strategies will need to be developed in most classrooms to assess certain process skills such as group contributions, and to enable formative assessment to function effectively (Black, 2001). As a priority teachers need a strong understanding of development in process skills in order to identify the focus for further learning. Hodson (1992) believes holistic assessment, where students engage in authentic science and experience constructive criticism from an expert teacher, has a powerful educative function because students develop the ability to appraise the quality of their own work. According to Harlen (1999) to understand how to take those next learning steps students need to be involved in the formative assessment process as much as possible. Teachers require strategies that focus attention on significant aspects of students' actions, such as observation of work, discussion of methods of inquiry and looking written work. Strategies such as peer review, and groups reporting to each other, enable teachers to hear students articulate their thinking, and students to gain the metacognitive skills for effective self-monitoring.

Sections 2.6 and 2.7 now look in detail at New Zealand developments in curriculum, pedagogy and assessment over the last 10-15 years, as they relate to the teaching and learning of scientific inquiry. The review examines the place of scientific inquiry in the SiNZC, teachers' pedagogical responses and the assessment of scientific inquiry-based learning within New Zealand's recently developed standards-based qualification system.

2.6 Science Curriculum Development in New Zealand

2.6.1 Influences on the Nature of the Science in the New Zealand Curriculum (SiNZC)

The current national policy statement in science, *Science in the New Zealand Curriculum* (SiNZC) dates back to 1993 (MoE, 1993). Bell et al. (1995) note that this curriculum development did not take place in a vacuum and that many New Zealand science educators had had experience, since the 1950s, of up to three national science curriculum developments in their teaching careers. Haigh (1995), in her account of the writing process, as coordinating writer for the curriculum document, sees the underpinning philosophy as co-constructivist (teachers and students together constructing meaning) in its pedagogy but acknowledges its overall eclectic nature. She claims the eclectic nature of the SiNZC reflects the consultative approach that was taken during the writing period and the informed debate that was an integral part of the process. Bell et al. (1995) contend, however, that the final result was not a negotiated document, one that has a theoretical integrity negotiated through debate, but rather one of compromise due to the tensions and different perspectives of the parties involved. They see it more like a 'marble cake' where different parts or aspects of the curriculum are coloured by different educational and political theories and ideologies. Of particular importance were the tensions and differences between the science educationalists and the views of the free-market advocates from the business sector. Bell et al. point to the mixed nature of the aims of science education in the SiNZC document as evidence of the unresolved thinking of these two groups. The aims are:

- science for all students and for future scientists,
- mixed theoretical perspectives on learning (neo-behaviourist and constructivist views of learning, Neyland, 1995),
- the tension between listing the skills and knowledge separately but wanting teachers to integrate them in the teaching and learning activities, and
- tension in the process and content debates.

Bell et al. (1995) make final comment that no one group has captured the official discourse of science education in New Zealand, although some outspoken critics of the curriculum saw the document captured by the ‘unsound’ epistemology of constructivism (Mathews, 1995; McMillan, 1995); this debate has since died away (Hipkins & Barker, 2002). It is worth noting that McMillan (1995) points out the term ‘constructivist’ does not actually appear in the document, and that there is no articulation of a theoretical perspective on learning and teaching. He sees this as a weakness of the document.

2.6.2 The structure and content of the Science in the New Zealand Curriculum (SiNZC)

The New Zealand Curriculum Framework (NZCF), an umbrella framework for the development of curricula in seven broad essential learning areas including science, determined the structure and parameters for the content of the SiNZC (Bell et al., 1995; MoE, 1993b). As prescribed in the NZCF, the writing team was required to create a structure for the curriculum statement comprised of Achievement Objectives and Assessment Examples over eight progressive levels and grouped in a number of strands (Haigh, 1995). The description of the essential learning area of science provided in the NZCF served to broadly define the content for SiNZC. Bell et al. (1995) summarise the key content for students to learn as:

- investigation and problem solving
- understanding scientific knowledge
- understanding the nature of science, and
- understanding the influence of science on people.

The NZCF also calls for each essential learning area to give scope to the development of essential skills (MoE, 1993b). Thus the SiNZC requires students to develop communication, numeracy, information, problem solving, self-management and competitive, social and cooperative, physical, and work and study skills as part of their science learning. Contexts for learning, teaching strategies and possible learning experiences were added after consideration by the writing team, but were not a requirement of the NZCF (Haigh, 1995).

There are six learning strands for SiNZC, four of which are known as the contextual strands, and the other two as the integrating strands (MoE, 1993a). The contextual strands are:

- making sense of the living world
- making sense of the physical world
- making sense of the material world
- making sense of the Planet Earth and beyond

The integrating strands are:

- making sense of the nature of science and its relationship to technology

- developing scientific skills and attitudes.

The curriculum writers found that the knowledge and concepts of science were relatively easy to place into a hierarchical list of learning outcomes, but had more difficulty achieving this in the skills and attitudes strand (Haigh, 1995). To encourage the teaching and learning of skills and attitudes in context and maintain a high profile for the development of scientific skills, a compromise was reached. The strand remained, but in the form of four broad learning progressions rather than eight. In addition, skills and attitudes were integrated into the content of achievement aims and objectives of the other strands. In the structure of SiNZC then the goals, concepts and skills of the curriculum are not separate and together they form the content of the curriculum. They are integrated into demonstrable learning outcomes that specify the achievement expectations at various stages of schooling. Like the curriculum statements of many other countries with the trend for behavioural outcomes, the SiNZC provides prescribed content for teachers to teach and students to learn (Crooks, 2002b; Orpwood, 2001).

The introduction to SiNZC promotes the view that scientific activity and scientific knowledge are interdependent: "Science is both a process of enquiry and a body of knowledge; it is an integrated discipline. The development of scientific skills and attitudes is inextricably linked to the development of ideas in science." (MoE, 1993, p. 14). The curriculum policy statement requires that content and process skills are to be given careful consideration and attention in teaching, with practical work, in particular investigations, cited as a vehicle for developing scientific understanding.

> Investigations provide key opportunities for students to extend their understanding in science. They also enable students to develop the scientific skills and attitudes required to enhance their ability to explore phenomena and events and to solve problems. It can be expected that, as they learn, students will show an increasing sophistication in the skills they use in their investigations. (MoE, 1993, p. 42)

The subsequent integration of skills and attitudes into the content achievement aims and objectives certainly supports this integrated premise, but whether the rest of the document is aligned fully with this principle is questionable. Elsewhere in the document the linear approach to science skills, as depicted in the 'Developing Scientific Skills and Attitudes' strand, implies generalisable, transferable, process skills that are independent of context. Hipkins and Barker (2002) fear that simplifying what is a complex set of processes down to a more potentially understandable linear form may, in the case of the SiNZC, promote a narrow interpretation of scientific inquiry (perhaps unintentionally). This portrayal could perpetuate the notion of 'the scientific method' and there is a danger that scientific investigation becomes associated exclusively with 'fair testing'. Such an interpretation would not in their view reflect the diversity of actual scientific practice, and Hipkins and Barker express concern about the effects of this view on classroom teaching and learning outcomes. They reiterate the need to achieve inquiry outcomes in tandem with content/theory where teachers "simulate as realistically as possible the things that scientists actually do when building knowledge of such products in the first place" (p. 23).

There is some evidence in the literature, based on teacher-reports and observations in classrooms, that in New Zealand classrooms teachers are offering more open investigative opportunities to students (Baker, 1999; Davies & France, 2000; ERO, 1996, 2000; Hipkins & Hooker, 2002), but there is little classroom based research that indicates the nature of any learning occurring. International and national measures of New Zealand student achievement, carried out during the last 10 years, do give some indications of student learning in investigative science. While there is some debate over the validity of measures made by the Third International Mathematics and Science Study [TIMSS] (MoE, 2002c) and Programme in International Student Assessment [PISA] (2000) studies, the results do suggest that New Zealand students have relative strengths in scientific inquiry (Baker & Jones, 2005). However, the PISA study (done with Year 5 and Year 9 students) did find that students' ability to plan and give scientific explanations for the findings of their investigations needed improvement. The achievement of the lower quartile students was of concern. The results of the National Education Monitoring Project [NEMP] (Crooks & Flockton, 2000a; 2000b; 2004), a national assessment programme conducted in New Zealand primary schools, found that students in science (Year 4 and Year 8) were as capable with open-ended investigative activities as they were with structured practical tasks.

2.7 Assessment and Qualification Reform in New Zealand – Some Implications for Student Learning

The large changes in the structure of school curricula in New Zealand over the last 10 years have been accompanied by correspondingly large modifications to the assessment systems (Crooks, 2002b). Change towards a criterion-referenced assessment system to accommodate the aims of an outcomes-based curriculum has been widely supported in principle, but indications are that there are challenges for teachers and students, particularly in the early days of implementation of new qualifications based on this system.

2.7.1 The National Certificate of Educational Achievement (NCEA)

Assessment reform in secondary education in New Zealand has been strongly influenced in recent years by the introduction of a qualification known as the National Certificate of Educational Achievement (NCEA). This prominence is reflected in the rhetoric (persuasive oratory and discourse) of forums such as the national science teachers' conference SCICON (New Zealand Association of Science Educators [NZASE], 2003) and various media reports (e.g., Blundell, 2003; Welch, 2003). Since accreditation in NCEA comes by way of standards (achievement and unit), based on their writers' interpretations of the New Zealand curriculum statements, this qualification brings in another curriculum interpretation which impacts on classroom practice (Carr et a., 2001; English, 1997; Knapp, 2002; Spillane, 2004).

Like the SiNZC, the NCEA qualification did not develop in a vacuum. It is valuable at this point in the review to recount the events that lead to the development of the NCEA. For over 30 years leading up to the introduction of NCEA in 1996 in its original form, existing procedures like annual external examinations, prescribed national courses and norm-referenced assessment results had been questioned (see, e.g., Lennox, 1995). Opponents to the former examinations-based assessment system asserted the qualifications did not fit the existing curriculum, and did not meet the needs of users or the full range of students (Strachan, 2001). Many authors believed the lack of validity of the examination system, the constraints on curriculum coverage and student learning styles, and the de-motivating effects for students of being labelled successes or failures were all seen to exert a damaging effect on teaching and learning (Lennox, 1995; Strachan, 2001). Standards-based assessment (in the form of achievement-based assessment) was widely practised in many schools in their senior classes, but final reporting within national awards was norm-referenced. During the long lead in period to assessment reforms, the advocacy for internal assessment and standards-based assessment grew in the belief that such assessment carried out by teachers would allow more flexible and relevant education, and more detailed and accurate information reported about students (Lennox, 1995). Reforms were delayed because of factors like inertia and resistance to change, politics and lobby group influences (Strachan, 2001), but in a restructuring of education administration in New Zealand during late 1980s the government of the day decided to bring about the necessary assessment reforms though legislation. These reforms signalled a new assessment culture in New Zealand (Crooks, 2002; Lennox, 1995; Scott, 2000) when assumptions underlying the existing national assessment and qualifications procedures were challenged.

In July 1990, the New Zealand Qualifications Authority was formally established under the Education Amendment Act (Crooks, 2002; Hattie, 2002; Lee & Lee, 2001). The Authority was legally empowered to oversee the development and review of national standards for all qualifications in post-compulsory education and training. A framework was to be established for administering national qualifications to increase the coherence of qualification systems so there were sensible and varied pathways towards appropriate qualifications for as many students as possible (Hattie, 2002).

Skills and knowledge gained in one context could be credited where required in any other context. A single qualifications framework, known as the National Qualification Framework (NQF), that contained provision for academic and vocational qualifications was adopted in 1991 (Crooks, 2002b). Qualifications under this system were to acknowledge a more diverse range of learning outcomes than traditional written examinations (Hipkins & Vaughan, 2002). The National Qualification Framework (NQF) was designed as a comprehensive quality control system incorporating an outcomes-based model derived from the NZCF (Crooks, 2002b; Hipkins & Vaughan, 2002,).

In education and other fields such as industry and commerce, ‘standards’ had become the catch cry, and many education reformers in the 1980s and the 1990s promoted outcomes-based models for curricula (Strachan, 2001). In New Zealand national curriculum development of the early 1990s was based on an outcomes model, and an assessment system that was in harmony with this model was clearly desirable. Assessment against standards and criteria was seen to complement the curriculum and learning by providing an appropriate and valid match between assessment methods and intended learning. In 1992 all qualifications on the NQF became based on a single component known as the *unit standard*. Unit standards represented nationally agreed levels of performance in the form of one or more competencies with accompanying performance-based elements and relevant performance criteria (Hipkins & Vaughan, 2002; NZQA, 1997). The units were essentially prescriptive behavioural statements about separate blocks of work within a given qualification that were transferable and able to be credited to another qualification (Lee & Lee, 2001). The NCEA was introduced in 1996 alongside other existing qualifications on the

Framework (Lee & Lee, 2001) and was based on unit standards. The traditional distinction between academic and vocational courses was discarded in favour of a more integrated 'seamless' approach favoured by the then Minister of Education. The unit standards would provide qualification users, like employers, information on the specific capabilities of qualification holders rather than just aggregated scores (Strachan, 2001).

Changes to the New Zealand assessment culture in the 1990s subsequently included: assessment against standards; assessment of a wider rage of attributes than had been possible through written tests or single examinations; and, the opportunity to assess for credit in a wider range of contexts than just the classroom (NZQA, 1996). As assessors in accredited organisations, teachers were encouraged, within the NQF, to integrate assessment with learning to suit particular aims, priorities and styles of learning. The assessment evidence could come from a variety of sources, including examinations and tests but the assessment activities had to be designed to produce true and valid assessment data by being consistent with the nature of the learning being assessed. Evidence could be gathered during normal everyday learning activities and more than one sample of evidence collected, for example, samples of work for a portfolio. Student performance was compared with the criteria within the unit standard and credit only given for successful achievement of all criteria. Assessment was to be authentic and embedded in the learning, and students given opportunities to try again if they had not achieved the required level of performance.

The changes required to bring the existing assessment practice in schools in line with the new reforms were major. The transition when it came in the early 1990s was rapid and huge for many schools and their communities (Strachan, 2001). There was no provision for the new assessment culture to evolve through incremental changes over an extended period, and the rhetoric of the NZQA failed to convince all, especially those who perceived they would be disadvantaged by the changes, that the new qualification was a viable proposition (Crooks, 2002b; Lee & Lee, 2001). Predictably, the introduction of unit standards in schools did not go smoothly.

Criticism of unit standards centred on:

- the reductionistic and atomistic philosophy of the NQF designers that all knowledge and skills can be reduced to predetermined sets of assessable competencies, and that competency-based assessment was the most valid assessment method
- the excessive fragmentation of subject areas into discrete ‘units’ of learning
- the ‘one-size-fits-all’ approach that assessed behaviour rather than underlying knowledge and understanding, and failed to take account of the needs of an increasingly diverse student population
- the pass/fail judgement that did not recognise excellence, and
- the manageability of the system because of the workload impact on teachers as assessors in schools and on the moderators. Unit standards were totally internally assessed so teachers were to make all professional judgements about students' achievements and stringent moderation became an issue especially interschool comparability (Hipkins & Vaughan, 2002; Lee & Lee, 2001).

By 1996 discontent with the nature of the NQF and unit standards had grown to such an extent that the government of the day decided to consult widely and review the direction in which the Qualifications Framework was headed (Lee & Lee, 2001). In 1997 a Green Paper (MoE, 1997) was released, acknowledging the limitations of the unit standard model and recommending a number of proposals for the further development of the NQF. These proposals included that:

- all major types of qualifications at all levels and across all subject areas regardless of how they were designed, taught or assessed should be registered on the NQF providing they meet or exceed a clearly specified quality benchmark
- the common currency of all qualifications be clearly stated outcomes or statements about what students know and can do
- a scale of nationally recognised excellence be integrated into the NQF for school subjects assessed against unit standards, and
- all existing national secondary school examinations be registered on the framework.

The Green Paper reiterated the overall goal of the NQF as ensuring that all major qualifications awarded in New Zealand were valued and credible in the eyes of students and employers. Qualifications were to provide recognition of learning that had taken place, skills that had been acquired, and a standard achieved.

In response to the recommendations made in the Green paper, a modified NCEA qualification was to take centre-stage and subsume other qualifications. This new form of NCEA adopted the *achievement standard* as the essential building block of the new Framework in the school sector, with unit standards having a higher profile in vocational courses both in and outside of schools. (Crooks, 2002b). The first stage of this modified NCEA implementation occurred in 2002 at the Year 11 level of secondary schooling (i.e., Level 1 of NCEA). Today students can gain the NCEA by accumulating credits by demonstrating they have met or exceeded the predefined outcomes of the achievement and/or unit standards (NZQA, 2001). Achievement standards, like the unit standards, were developed from national curriculum statements and existing examination prescriptions (Hipkins & Vaughan, 2002). Whereas unit standards describe what students are able to know and do using precise performance criteria in order to gain credit at one level of achievement, achievement standards judge student performance for credit at three levels (achievement, achievement with merit and achievement with excellence) using less specific criteria than those used for unit standards. The three levels of achievement in achievement standards broadly correspond to progressively higher levels of thinking, such as those levels depicted in various classification systems of the cognitive domain like Bloom's Taxonomy (Anderson & Krathwohl, 2001; Bloom et al., 1956). A student might be typically expected to 'describe' at achievement level, 'explain' at merit level and 'apply' or 'evaluate' at excellence level. These differentiated results are intended to give parents, tertiary institutions and employers a more comprehensive and useful picture of student achievement (NZQA, 2001). While the largely vocationally-related unit standards are based on internal assessment, at least half the academic achievement standards are to be externally assessed to address critics' concerns about issues to do with internal assessment, such as moderation and teacher workload (Lee & Lee, 2001).

The government, through the MoE and NZQA, provided a number of support activities to assist schools and teachers to implement NCEA (ERO, 2004). They included a mix of professional development support and resources. NZQA school relationship managers were appointed, primarily to provide systems advice to teachers and visit schools to share good practice. Professional development days were contracted to the MoE, resulting in a nation-wide professional development project on a scale never seen before in New Zealand (NZQA, 2000). This massive undertaking, involving most secondary teachers, occurred over a four-year period and ended in 2003. Education advisory services continue to provide support on request, but ongoing training has largely ceased. Resources for teachers (and students) included exemplars of assessment tasks and marking schedules for internally assessed achievement standards that were available for downloading from an MoE website. There was a commitment by NZQA to the professional development of teachers prior to implementation of NCEA.

Controversy still surrounds the NCEA qualification. The qualification represents massive and ambitious reforms in assessment and curriculum delivery, including ongoing complex design and development (Black, 2001). Many continue to express concern about the appropriateness of using a standards-based system in a high stakes qualification like NCEA because of the inherent problems associated with specification of clear standards and moderation (Black, 2002; Elley, 2003; Lee & Lee, 2001; Welch, 2004). Crooks (2002) worries in this new climate of accountability about the excessive precision required of teachers in their assessment of student performance. Manageability in terms of teacher workload, and inadequate resourcing to support the process of implementation continue to be issues for many schools, and the teaching profession in these early stages (Blundell, 2003, Hipkins & Vaughan, 2002).

On the positive side, early indications are that the NCEA is already improving access to qualifications for a wider range of students than traditional qualifications centred solely around examinations (Hipkins & Vaughan, 2002, 2005; Hipkins et al., 2004). A sizeable proportion of the standards can be assessed in the context of normal classroom work. Internal assessment activities, integrated into the teaching and learning, have the potential to give more valid and useful indications of a wider range of student abilities and attitudes than traditional external methods. This is particularly so in the case of practical work, performances and ongoing investigative or research work in subjects like science. There are some indications that students place slightly more value on credits gained from internal assessments (Hipkins & Vaughan, 2005).

With its similarities to the Queensland assessment model (Black, 2003a), in its criterion or standards based assessment framework, internally assessed components and panels of moderators, NCEA may also have to address some of the formative-summative tensions reported in the literature. Policy supporting NCEA recognises the powerful influence of assessment on teaching and learning, resulting in a qualification design that deliberately places a more explicit and transparent focus on the outcomes students can be reasonably expected to achieve (Fancy, 2001; Strachan, 2001). The stage could be set for enhanced learning by giving teachers and students improved access to the tool of formative assessment.

Some possible positive and negative consequences of the introduction of NCEA are signalled in the findings of a study involving 18 teachers from very different secondary schools that sought to explore relationships between the change to the standards-based NCEA qualification regime and associated changes in Year 11 science and mathematics teaching (Hipkins, 2004). Since no recent, large scale, studies had been done of actual classroom practice in New Zealand schools of science teaching, there were no available base-line data against which change could be monitored in this study. Consequently, the Hipkins study relied on teacher's own perceptions of change in classroom practice and student achievement since the introduction of NCEA. Despite the common misgivings of teachers about the impact of NCEA on classroom practice and achievement, and the tendency for teachers to overlook small incremental changes, a number of potentially positive changes can be identified. Significant increased practice was reported in:

- assessment of a range of levels with increased use of higher order tasks such as application, analysis, and synthesis of ideas
- learning for meaning rather than for content coverage
- developing understandings of use of language conventions in science, and
- the use of new technologies in assessment.

In interviews the teachers in Hipkins' (2004) study revealed that the need for students to demonstrate higher order thinking skills to achieve merit and excellence levels prompted them to give greater assistance to their more able students. Classroom learning and assessment tasks often had to be adapted or new ones created to help students develop these higher order skills. Interestingly, in the externally assessed standards not all students were being given the opportunities to acquire these critical thinking skills, only those deemed more able. However, this was not the trend in the internally assessed practical investigations. Teachers talked about a change in emphasis rather than a new implementation, by purposefully using more open and less structured teaching and learning episodes to give students opportunities to develop higher order investigative skills. One teacher reported students enjoying this change to a less structured approach, and viewing it more as doing 'real science'.

The Hipkins (2004) study also noted that these changes in emphasis in teaching and learning were counterbalanced by decreases in some forms of practice. For example, teachers used fewer strategies to develop students' metacognitive awareness, that is, teaching students to think critically about their learning. In one instance, this was because of greater emphasis placed on teaching content detail in the belief that this would assist students to achieve the merit and excellence levels. In another change, adherence by some teachers to the reporting schedule framework developed for the assessment of the achievement standard by NZQA sources, had a narrowing effect on the nature of the science investigations able to be undertaken. Several other teachers encouraged their students to work outside the constraints of the formal template by writing less structured reports.

As a rule the teachers were not assessing students' achievements for the standard on the basis of those abilities demonstrated during the teaching and learning programme (Hipkins, 2004). Assessment, including formative assessment, was typically done under 'examination' conditions. Formative assessment was not seen as an integral part of teaching and learning but rather a separate process. For many of the teachers formative assessment equated to practice sessions where the focus was on "assessment of assessment", in other words, how successfully they were meeting assessment requirements rather than the learning of scientific skills and knowledge. A few teachers reported using strategies to deepen students' understanding, and in one instance a strategy that specifically helped students develop their ability to self-monitor and extend their learning. This suggests formative assessment can sit within the NCEA structure successfully, but Hipkins (2004) makes the point that teachers need strong pedagogical knowledge of formative assessment practice, and belief in its value, for widespread teacher change to occur. Implicit in this is teachers' need for expertise about the nature of scientific processes (Sadler, 1989), and relevant pedagogical content.

2.7.2 NCEA assessment of scientific investigation

As a policy statement the SiNZC couches its message in broad terms and as such does not specify the procedural details students are required to acquire and demonstrate in practical investigative work. In the context of Year 11 schooling, the initial year of access to qualifications for most secondary students in New Zealand, the first interpretation of the inquiry intent of the SiNZC for a qualification was made in the School Certificate Science prescription. The original authors of this prescription, in the view of Hipkins and Barker (2002), gave practical investigation a rather superficial treatment. For example, there was no requirement in the prescription for students to be assessed on their ability to design and carry out experiments; as a consequence teachers tended not to emphasise these particular skills in their teaching programmes. This reflects similar overseas experience (Atkin & Black, 2003; Hudson, 1992).

In contrast the Level 1 Unit Standard *Carrying out a scientific practical investigation with direction,* the first standard for Year 11 students in investigative science under NCEA, required a fuller, in-depth treatment because of its greater specificity of learning outcomes. What became quickly apparent were weaknesses in many secondary teachers' understanding of aspects of scientific investigation (English, 1997). These weaknesses were exposed because the new approaches called for a broader range of knowledge and skills to be assessed. Teachers were required to reinterpret the intent of the SiNZC by the modifications undergone at the qualifications site of influence. English reports that moderators, working with teachers accustomed to the School Certificate science prescription and assessing this unit standard for the first time, found few teachers submitted valid assessment activities. It appeared teachers lacked deep understanding of the various aspects of a science investigation required to assess them. The moderation thus became a source of professional development for teachers increasing their understanding of how science process is integrated with science concepts, and how to identify appropriate means of assessment. In research supporting the use of standards in New Zealand for the teaching and assessing of procedural knowledge in practical work (Foster, 2000), it is significant that a linear methodology for science investigation continues to be promoted.

The Level 1 Unit Standard *Carrying out a scientific practical investigation with direction* was effectively replaced by the very similarly named Level 1 Achievement Standard, Science 1.1 *Carrying out a practical science investigation with direction* (AS1.1) in 2002, when achievement standards were introduced into NCEA (Hipkins & Vaughan, 2002). The standard contains considerable detail for teachers and students about the aspects of learning to be demonstrated for accreditation at the three levels of achievement (MoE, 2003, see Appendix A). The standard is further specified by the required use of student procedural instructions for the investigation and templates for planning and reporting by NZQA. These are provided via exemplars made available on a MoE website (see Appendix B for example "Bubble Trouble").

Science Achievement Standard 1.1 states that student investigations should be based on situations arising from content, drawn from content up to Level 6 of the SiNZC, or the Te Tauaki Marautanga Putaiao: He tauira (National Science Curriculum Statement in Māori), (MoE, 1994b). The standard defines an investigation as an activity covering the complete process from planning to reporting, and is to involve the student in the gathering of primary data. Under direction from the teacher, students are expected to:

- produce a workable plan, containing a purpose, provision and evidence of trialling, key variables and how they will be controlled, a method for data collection and consideration of factors such as sampling, bias, sources of error and sufficiency of data
- execute the plan, collect appropriate data and record in a table or other systematic way, and process to establish a relevant pattern or trend. Data processing is expected to usually involve calculations such as averaging, and
- interpret the processed data in relation to the purpose of the investigation and write a report following written guidelines from the assessor. Sections of this report are to usually include the purpose and final method used, recorded and processed data showing links, interpretations, a conclusion linking findings to the purpose, and an evaluation or discussion.

To assist teachers' interpretations of this standard prior to implementation, as part of the government support initiatives for NCEA, working groups developed a number of science assessment tasks, including assessment schedules and exemplars of graded student work. These materials were made available to teachers (and students) on a MoE website. However, many tasks were not trialled and there were a number of teething problems with the workability of these tasks in practice (Hipkins et al., 2004). After initial use by teachers and students Hipkins et al. feel many of the developed assessment tasks do in fact attempt to be authentic extended investigations, but that the tasks need refinement.

Hipkins and Booker (2002) describe an attempt by a science teacher to enable her students to experience authentic investigations as they worked towards AS1.1. This teacher sought to simulate 'real science' conditions by embedding the investigation, as scientists would, in a familiar theoretical field. For students this meant gaining prior knowledge in the science context of the investigation, which was to be the properties of light. Like scientists, the teacher also wanted students to collaborate in planning, trialling and modifying. Time was given for students to carry out preliminary practicals into the behaviour of light, and to consider how the results and conclusions of these exploratory activities could be used to answer various questions. Students were given the opportunity to use the suggested template (provided by NZQA) for AS1.1, play with equipment, prepare draft plans for new investigations, trial these plans to see if they were workable, and finally carry out another familiar investigation into light. The teacher gave students feedback on their performance, but no grade. The summative assessment was a modified version of a task supplied by an

MoE website, containing extra information provided by the teacher that was pertinent to the investigation. The context was again the properties of light, and students performed the investigations individually. Hipkins and Booker (2002) did not comment on the nature of the learning outcomes for these students in this paper.

In addition to the materials produced by MoE working groups to assist teachers interpret Science Achievement Standard 1.1, texts related to the achievement standard were quickly produced by various publishers. They ranged from textbooks providing content and exemplars (e.g., Hannay et al., 2004), to student laboratory manuals (e.g., Cooper, Hume, & Abbott, 2002). Most tend to closely follow the format and requirements of the standard and accompanying exemplars in their interpretation of carrying out scientific investigations.

2.8 Concluding Thoughts

This literature review reveals a number of points worth pursuing. Of particular interest are certain context features that have the potential to impact on the nature of the student-experienced curriculum. At this stage it is timely to revisit the research questions in turn and highlight key findings from the literature review that are pertinent to each question.

The first question seeks to determine *what* science New Zealand students are learning in NCEA classroom programmes for Science Achievement Standard 1.1 *Carrying out a practical investigation in science with direction.* On examination the intended curriculum, as depicted in the SiNZC policy statement, appears to be sending out mixed messages about the nature of scientific investigation. On one hand SiNZC acknowledges the interdependency of scientific activity and science knowledge, but on the other it appears to promote a more narrow, linear view of scientific inquiry. Research suggests this eclectic approach leaves the policy statement open to various interpretations and the possibility that different students may experience different implemented curricula.

While the intended curriculum may provide guidelines for classroom practice, this review also suggests that it is unlikely students will receive this curriculum in the intended form. This has implications for another of the research questions, which seeks to determine the degree of match between the intended curriculum and the operational curriculum. Research suggests that various sites of influence within the context of this study, such as the Science Achievement Standard 1.1 *Carrying out a practical investigation in science with direction*, government agencies, resources, school assessment systems and classroom teachers views on teaching and learning are likely to have impact on the intended curriculum and influence the nature of the student-experienced curriculum. Perhaps links can be drawn between the nature of the student-experienced curriculum and the effect of these sites of influence. While a number of potential influences have been identified in this literature review, this investigation could reveal more. Consideration therefore needs to be given to research methodologies and methods that will reveal such influences, if they do in fact exist.

Another point of interest is how student learning in this educational context matches learning currently judged ‘worthwhile’ as defined in the literature. Is the learning the result of open-ended inquiry, as advocated by progressive thinking in science education, or is learning shaped by the more traditional, closed, recipe-style practicals?

Whatever the nature of the actual learning occurring in the student-experienced curriculum, the remaining research question is concerned with finding out *why* and *how* this particular learning is occurring. Can the factors and conditions that result in this learning be clearly identified? If 'worthwhile' learning is not occurring, what are the reasons? Perhaps students are learning alternative concepts and skills, or maybe learning is being limited by certain factors. Again careful thought needs to be given to research approaches and design that will yield useful, meaningful answers. Such answers need to furnish feedback and feedforward information that other researchers and practitioners can utilise to improve learning outcomes for students. These answers lie in investigating classroom realities, observing teachers and students at work and seeking students' perspectives on what they are learning.

2.9 Summary

This chapter has provided an extensive account and critique of the literature and events relevant to the research questions under consideration in this study, in recognition of the multitude of factors that have the potential to impact on the curriculum students experience in classrooms. The following chapter describes the methodology, and research methods, techniques and tool used in the study, and provides justification for decisions in the research design process related to the trustworthiness of the study.

CHAPTER THREE

CHAPTER THREE

METHODOLOGY

Overview

This chapter provides an account of the methodology used in this inquiry. The first section summarises the paradigms and associated methodological approaches that are used in educational inquires, and the rationale behind the decision to use an interpretive framework to guide this study. Following sections include the consideration given to the trustworthiness of this inquiry, and a detailed description of the research design including the techniques and tools used for data collection and analysis. The chapter concludes with a discussion of ethical issues and details how they were addressed in the study.

3.1 The Theoretical Framework that Guided this Inquiry

In offering advice to new researchers, Zevenberg and Begg (1999) stress the importance of identifying the theoretical perspective used to guide the approach to data collection and analysis. This theoretical framework, or paradigm (Kuhn, 1962, p. x) which the researcher draws upon as a basis for the study, is the pervading theme of the project. Patton (1990) talks of such a framework being strategic because it provides basic direction for the research. The paradigm is a framework for action giving the research inquiry an internal coherence because it determines what is looked for, the way in which observations are construed and how emerging problems are addressed (Coll, 2002; Gipps, 1994). Each paradigm is characterised by particular methodologies, that is, research methods and techniques that have been developed for the problematics of that paradigm. Some writers suggest that researchers adhere to the methods that are identifiable with certain paradigms (Potter, 1998) while others suggest selection of methods should be guided by “fitness for purpose” (Cohen et al., 2000, p. 47).

Education is a social activity where humans engage in teaching and learning about their world. Educational research therefore sits naturally under the wider umbrella of social sciences research, and is influenced by the trends and approaches characterising this field of endeavour. There are three research paradigms, or lenses through which research practice is commonly examined, that broadly describe the approaches taken to research in the social sciences (Cohen et al., 2000). They are:

- scientific and positivistic paradigms
- naturalistic and interpretive paradigms, and
- critical theory paradigms.

Educational research is concerned with developing a knowledge base around the issues of education, and finding solutions to the problems of teaching and learning. The advice of Shulman (1981) to researchers, in selecting a particular paradigm and methodology for a study, is “first understand our problem, and decide what questions we are asking, then select the mode of *disciplined inquiry* most appropriate to those questions” (p. 12, original emphasis).

3.1.1 Scientific and positivistic paradigms

First used by the French philosopher Auguste Comte (1798 –1857), the term positivism today refers to a paradigm that accepts natural science as the model for acquiring and establishing human knowledge (Beck, 1979; Bullock & Trombley, 2000; Burns, 1994; Lather, 1992). The paradigm takes a normative view of human behaviour, believing actions are rule-governed, and should be investigated by the methods of natural science. In their ontology (theory of existence), positivist educational researchers operate from the realist position of an objective reality and are unobtrusive, objective observers of their research subject matter. Since their epistemology (philosophical theory of knowledge) views all true knowledge as scientific, positivist education researchers adopt what is termed a nomothetic approach to their research (Burns, 1994), seeking to discover general and universal laws using the methods and techniques of natural science, hence the interchangeable use of the terms scientific and positivistic in the literature. Postivistic approaches to education research tend to be associated with behaviourist views of learning (Hipkins et al., 2001)

Science has fundamental tenets that have direct bearing on the nature of scientific inquiry (Cohen et al., 2000). These underlying assumptions are:

- determinism - the belief that all events have causes and if these causal links can be discovered, then the world can be better understood, controlled and predicted
- empiricism - the view that knowledge can only be considered reliable and justifiable if it is derived from evidence gained directly through experience. It must be verifiable
- parsimony - the expectation that all scientific explanations of phenomena and expressions of ideas be as economical as possible, and
- generality - the creation of abstract generalisations for the world at large from observations of the particular using the inductive–deductive methods of reasoning.

The ultimate goal of scientific research is theory, which Kerlinger (1970) defines as "…a set of interrelated constructs (concepts), definitions, and propositions that present a systematic view of phenomena by specifying relations among variables, with the purpose of explaining and predicting the phenomena" (p. 9). A theory's strength lies in its potential to give direction to new research and discoveries – its predictive power. It is through a theory's hypotheses that much research proceeds (Cohen et al., 2000). These conjectural statements (preconceptions of what might be true based on the premises of the theory) allow investigations of relationships between variables and between cause and effect. To be useful, a theory must have wider applicability than just explanation. It must provide the capability to identify gaps in knowledge and generate new ideas and questions.

A positivist researcher in education concentrates on the objective world and tries to exclude the subjective world as much as possible. The emphasis is on control and replication. Often described as technist, positivistic methods typically use experimental and quantitative modes of inquiry (Cohen et al, 2000). Observation, measurement, classification, the careful manipulation of variables in attempts to identify relationships, and mathematical analysis to establish general laws and principles, are common features of these methods (Coll, 2002; Kennedy, 1997; Peterson, 1998; Potter, 1998). Surveys and questionnaires are probably the most commonly used research instruments in positivistic-based education research.

Positivism as a research paradigm in social sciences was dominant until some 30 years ago when its popularity began to wane. Many social science researchers felt that the paradigm was not yielding the depth and body of knowledge hoped for, and increasingly scientific methodology was seen as problematic when applied to the complexities of human behaviour (Kennedy, 1997; Lather, 1992). These researchers have come to believe that humans' actions simply do not occur in orderly and regular ways like phenomena in the natural world, and instead see human society as highly complex. Traditional scientific methods are now increasingly seen as inadequate, inappropriate, frustrating, and non-profitable in social sciences research. They "have not proved successful in providing major advances in the understanding of human life" (Potter, 1998, p. 118).

This wave of dissatisfaction with the positivistic-based research in the social sciences was also being reflected in the educational world (Kennedy, 1997). As a 'scientific' social science 30 years ago, educational theory was still in its early stages (Cohen et al., 2000), despite a century of work since its beginnings as a form of educational psychology influenced largely by behaviourism (Eisner, 1983). Much of the collected data was descriptive and there were no grand theories on which to predict human behaviour or base future research. Furthermore, the research seemingly had little impact on classroom practice (Atkin & Black, 2003; Kennedy, 1997; Woolnough, 2001). This has led to a growing awareness from members of all the social sciences, including those in education, that different perspectives and methodologies were needed to better understand human dynamics (Cole & Engestrom, 1993; Kennedy, 1997; Lather, 1992; Peterson, 1998; Walshaw, 2001). There is a belief that educational research "may have to encompass broader views of what counts as scholarly inquiry" (Atkin & Black, 2003, p. 128) to help bridge the gap between research findings and education practice. So began the anti-positivist school of thought whose members today have many and varied approaches to this central theme. A great array of terms is used to describe the methods and approaches of researchers with an anti-positivist persuasion. Some of the most widely used terms include interpretive, naturalistic, constructivist, ethnographic, emancipatory, critical, feminist, and deconstructionist (Lather, 1992).

All in the anti-positivist movement call into question the validity of the positivist world-view (Lather, 1992), and its notions of truth and legitimate knowledge. The consensus amongst such thinkers is that reality is subjective (Eisner, 1983), and that objects of thought are simply words – they do not have a separate existence (Cohen et al., 2000). Antipositivists' ontological position is thus nominalist, as opposed to realist. Epistemologically, anti-positivists take the stance that knowledge has to be personally experienced and social reality is the result of individual cognition and interpretation. The "focus is on the constructed versus found worlds" (Lather, 1992, p. 89). Learning theories that accept this premise of knowledge generation include constructivist, socio-cultural and linguistic views of learning.

3.1.2 Naturalistic and interpretive paradigms

In contrast to the scientific-positivistic paradigm, the interpretive paradigm asserts that general universal laws do not govern human behaviour and understanding of this behaviour is not gained by using the methods of natural science (Cohen et al., 2000).

Interpretivists argue that humans, unlike inanimate objects, are autonomous and exercise agency, demonstrating intention and choice in their actions. People also have the unique ability to interpret their experiences and to give meaning and representation to them. As self-aware beings capable of thought and with the power of language, humans construct theories about their world as they attempt to make sense of their surroundings (Guba & Lincoln, 1989). Applying the laws of natural science for "grasping the essential nature of human experience and behaviour" (Cole & Engstrom, 1993, p. 43) is therefore considered inadequate - what is needed is an understanding of social reality, as individual observers perceive it (Guba & Lincoln, 1989).

Such understanding of social reality can be developed in an idiographic approach, which has its roots in ethnographic studies (Burns, 1994; Coll, 2002; Goetz & DeCompte, 1984; Potter, 1998). In an idiographic approach, knowledge of the situation is gained from the viewpoint of individuals who take part in the ongoing action being studied. These 'situated activities' are studied through the eyes of the participants, rather than those of the researcher, to ensure maximum fidelity, because human events and behaviours are strongly affected by context. The rationale is that deep insights into individuals' interpretations of their reality can only be gained if their social world is studied in its natural state with little interference or manipulation by the researcher, thus the term 'naturalistic' is often applied to such research methods (Patton, 1990). These methods "preserve the integrity of the situation in which they are employed" (Cohen et al., 2000, p. 26). The researcher takes on a participatory rather than observer role, characterised by the meticulous analysis of the participants' accounts of their behaviour or action in the particular social episode being studied. In education, for example, there is a growing school of thought that such in-depth analysis enables researchers to learn from practicing teachers about effective ways of improving students' achievement (Atkin & Black, 2003).

In the interpretive paradigm, empirical data is typically gathered using methodological approaches such as case studies, ethnographies and narratives (Kennedy, 1997), and qualitative inquiry methods or tools such as unobtrusive observation, participant observation and in-depth interviews (Coll, 2002). These techniques produce a wealth of information that allows the researcher to examine particular issues in detail and depth. Methods of analysis, such as grounded theory and discourse analysis (Potter, 1998), look for patterns and relationships between ideas or categories of things in order to generate theory. Theory in interpretive research is emergent, diverse, multi-layered and multifaceted. It does not start from preconceived hypotheses, but is rather a set of meanings derived from and 'grounded' in the data gathered during research (Glaser & Strauss, 1967). The researcher "paints a picture" (Bishop, 1997, p. 30) that allows the participant's voice to be heard, and for others in the research field to reflect upon in terms of applicability to their experiences and contexts (Haigh, 2000).

3.1.3 Critical theory paradigm

Proponents of the critical theory paradigm claim that not only positivist but also interpretivist investigations of social behaviour are inadequate, because they both neglect the political and ideological contexts in which research agendas are designed (Cohen et al., 2000). Critical theorists argue that the other paradigms accept rather than question given agendas for research, and therefore maintain the status quo for marginalised and disempowered groups in society. They believe in emancipatory research that is deliberately political and transformative in its intent. This approach seeks a more inclusive position (Walshaw, 2001), respecting the intellectual and political abilities of disenfranchised individuals and groups. The purpose of critical theory is to provide an understanding of human situations and phenomena so they can be changed to create a more egalitarian society. Issues of empowerment, diversity and equity are to the fore. Adherents to this paradigm (Harding, 1987; Lather, 1992), including feminist and race-specific workers, have a firm view on what behaviour in a democratic society should be. They are deliberately political in seeking to change what they see as oppressive social conditions (Walshaw, 2001).

This critical theory perspective is gaining supporters in educational research, and possibilities for developing emancipatory pedagogies and practices are being explored. Central to this is the recognition and valuing of difference (Walshaw, 2001), and the validating of different knowledge bases. These require new theoretical and analytical approaches such as ideological critique (Cohen et al., 2000). This strongly practical, rather than theoretical, methodology seeks to identify the values and practices of particular dominant groups in a given 'problematic' situation and expose the vested interests at work. Using reflexive practice, the situation is first described and interpreted in what is termed a hermeneutic exercise to gain understanding of the human behaviour and social institutions involved. An analysis of the causes and purposes of the existing situation, and an evaluation of their legitimacy follow this. This second stage of the critique aims to reveal to the participants the ideologies at work and how they influence their views and practices. These same views and practices may be working to maintain a system that keeps participants disempowered and oppressed. The next step is to devise a strategy for altering the situation and enacting it. Finally, an evaluation of the resulting situation in practice is done.

A methodology suited to critical theory investigations is action research. This combination of research with action sees participants as collaborative researchers, studying problems in their everyday social practice using a form of disciplined self-reflective enquiry (Kemmis & McTaggart, 1988). The aim is improvement, with better understanding of their practice, by practitioners (Engstrom, Engstrom & Sunito, 2002). This methodology is intended to bridge the gap between research and practice. In education the reflective aspect of action research is viewed as a form of professional development for teachers (Keeves, 1998), and as a means of increasing the relevance of research for teachers (Kennedy, 1997). It is also advocated as a way of combating the perceived failure of research to change educational practice (Peterson, 1998). Other techniques increasingly used for critical inquiries include discourse analysis (Nikander, 1995; Cole & Engestrom, 1993) and narratives (Bishop, 1997; Burman, 2001).

It is worth mentioning at this point that Lather (1992) notes a fourth category of methodologies emerging that can be added to those described by Cohen et al. (2000). This fourth category includes what is termed the deconstructionist methodologies such as post-structural and post-modern. In essence these methodologies represent the move to challenge all assumptions, standards and canons of science, even to the concept of paradigm in research. Hence, they represent some of the features of critical theory, but take these to a more emphatic position.

3.1.4 The research goals for this study

The research reported in this thesis seeks to identify the nature of the science that students are actually learning in New Zealand classrooms today, and how this relates to the intentions of the Science in the New Zealand Curriculum (SiNZC) document. From the perspective of students in real classrooms the research aims to determine *what* science students learn, and *why* and *how* they learn that science. Guided by indications from the literature of gaps in the knowledge about the classroom curriculum in science, the research questions focus on the learning of investigative science within the setting of a new national qualification regime in New Zealand, known as the National Certificate of Educational Achievement (NCEA). Having decided my research aims and questions, and investigated the research paradigm options available to me it was time to follow Shulman's (1981) advice and select a paradigm and methodology best fitted to the purpose of this study. The decision was made to adopt an interpretivist approach. The reasons for this approach are now described.

3.1.5 The case for interpretivism as the paradigm of best fit

For this inquiry's purposes, a naturalistic or ethnographic approach appeared most appropriate because the inquiry concerned people's activities in their natural settings (Potter, 1998). As the researcher I hoped to get inside the understanding and culture of the group, by immersing myself in the situation, to comprehend how the group members perceive their world and why they act the way they do (Woolnough, 2001). Since this project's primary goals were to reveal participants' views of reality and how they engage in sense-making, the interpretivist paradigm seemed the best match (Ball, 1984; Boaler, 2000; Cohen et al., 2000; Cowie & Bell, 1999; Haigh, 2000; Guba & Lincoln, 1989). Holistic and interpretive research style were called for (Atkin & Black, 2003) because the "behaviour and, thereby, data are socially situated, context related, context dependent and context rich" (Cohen et al., 2000, p. 137).

Further justification for the interpretivist approach lay in the uncertain nature of the possible findings. It was the students' views that I was pursuing and I did not know in advance what I would find, nor was I sure what to look for although my research goals did provide broad areas of interest to guide my initial observations and data collection (Spindler & Spindler, 1992) – again features compatible with an interpretivist approach.

For my study to validly reflect participants' views and actions in their real world setting my stance as researcher needed to reflect the naturalistic approach of the inquiry. I was therefore required to declare my presence and describe what my personal, social, interactional and ethical position would be during the research (Graue & Walsh, 1999; Spindler & Spindler, 1992). The theory was to be emergent, grounded in the data, and I was prepared for the likelihood of the research objectives changing direction, focus and scope as the investigation proceeds (Ball, 1984). The interpretivist paradigm would allow the diverse, complex and unique classroom situations in which my research was likely to take place to be acknowledged and explored (France, 2001; Haigh, 2000). Finally, I believed the potential for my findings to have credibility for curriculum developers and teachers, and being integrated into practice, were greater too if I took an interpretivist view (Atkin & Black, 2003).

Since the study concerned social behaviour and my research goals sought findings that might improve educational practice, I also needed to consider whether a critical theory approach would serve my purpose (Cohen et al., 2000; Patton, 1990). I chose not to take this approach for several important reasons. First, while my principal aims were to determine the learning reality for students and the manner in which it occurred, I was not seeking to intentionally emancipate and empower participants as part of the research process. Secondly, I judged it inappropriate to employ research methods, such as action research, to facilitate participants' involvement as collaborative researchers in revealing the ideologies at work in the educational context. The promotion of participants' self-reflective practice may well improve educational outcomes for students, but such methods were unlikely to reveal the data required to achieve my research goals within the available time frame. Such methods also work against gaining insights into individuals' interpretations of their reality through minimal interference from the researcher.

With the guiding theoretical framework chosen I needed now to consider certain problematics that needed addressing within this paradigm, including those related to the trustworthiness of the research process.

3.2 Data Gathering Considerations for Trustworthiness

In interpretive inquiry, to faithfully reconstruct the reality of those being studied, the researcher needs to use methods of gathering evidence that are compatible with the humanistic nature of the inquiry. Appropriate methods include unobtrusive observations, interviews and document analysis. Once selected, these methods require careful scrutiny to ensure what Guba and Lincoln (1989) term the overall ‘trustworthiness’ of the research design. Trustworthiness refers to the goodness or quality of the research process and requires that issues such as credibility, comparability and transferability, dependability, confirmability, sampling and ethics be explored and addressed in the research design. The following sections explain how these issues contribute to the trustworthiness of an interpretive-based study and how they can be addressed.

3.2.1 Credibility

Credibility relates to the authenticity and worth of the study, and is judged by the closeness in match between the constructed realities of the researcher and those of the participants: “The demonstration that the researcher’s interpretations of data (the findings) are credible to those who provided the data” (Eisenhart & Howe, 1992, p. 651). Thus to assure high levels of credibility an interpretive study needs to employ methods of data collection and analysis that genuinely reflect the situation being studied, and the interpretations the participants have of their social reality. These requirements can be met by strategies such as prolonged engagement, triangulation, and respondent validation (Cohen et al., 2000; Guba & Lincoln, 1989, Patton, 1990).

Ultimately the study's real worth is measured by the insights it provides for the reader to do their job better and improve their practice (Woolnough, 2001). Greater confidence in the credibility of the study is also promoted by generative research that raises new questions and issues, contains convincing arguments, and includes a clear audit trail provided by the researcher (Potter, 1998).

3.2.2 Comparability and transferability

In interpretive research it is the reader of research findings who judges their degree of comparability and transferability (Guba & Lincoln, 1989), by deciding the pertinence and relevance of the study findings to other situations. Maximising the opportunities for readers to make these meaningful comparisons with other similar or dissimilar situations requires rich or thick description of the participants, the setting and circumstances – "a verbal picture" (Goetz & LeCompte, 1984, p. 238). Explicit accounts of the characteristics of each studied group and the analytic categories used in the research are also essential for the reader to draw parallels between the study, and the particulars of their own situation. Educational case studies, as a research approach, facilitate these sorts of comparisons (Cohen et al., 2000) and they are discussed later in this chapter.

If applicability to other situations is to be a key outcome of the research, then comparability and transferability obviously have to be well catered for in research design. Methods that enable the researcher to provide thick descriptions of the setting and actions, such as comprehensive documenting of the characteristics of the participants' actions and thoughts and classroom procedures, are required so readers can judge the relevance and worth of the findings (Boale, 2000, Lincoln & Guba, 1989). For example, to shed light on the way participants use language constructs to actively gain everyday meaning (Nikander, 1995), researchers can use interpretive-based analysis of data gleaned from students' thinking; teachers' views, planning, and delivery; and textual teaching and learning materials. Such analysis can also give insights into the nature and extent of the social practices that constitute teaching and learning in the classroom setting being studied, helping to build an 'inferential bridge' to those other groups to whom the findings may be applicable (Shulman, 1981).

3.2.3. Dependability and confirmability

Dependability and confirmability in interpretive-based methodologies include "fidelity to real life, context and situation-specificity, authenticity, comprehensiveness, detail, honesty, depth of response and meaningfulness to the respondents." (Cohen et al., 2000, p. 120). Consequently in qualitative inquiry observation is characteristically prolonged, extensive and often repetitive to establish the dependability or stability of the gathered data, that is, a close fit between what the researchers regard as data and what actually occurs in the situation being studied (Spindler & Spindler, 1992). Changes in methods and foci are integral features of interpretive studies and these changes must be duly considered when attempting to maintain the stability of data. An interpretivist researcher will therefore carefully track and record these changes to show the development of the inquiry process and the rationale for the changes. Careful auditing of such events helps to assure, not only the dependability of the data, but also its confirmability, that is, the assurance that the personal biases, values and motive of the researcher have not unduly influenced the results of the inquiry (Guba & Lincoln, 1989). Respondent validation (participant checking) of the raw data, and the processes used to collate and categorise the data, is a strategy frequently used in qualitative inquiry to enhance the dependability and confirmability of an interpretive study.

3.2.4 Selecting the research group

Deciding upon the research group involves defining and selecting the population on which the research will focus (Cohen et al., 2000). In interpretive research factors such as time, expense and accessibility often necessitate the researcher obtaining data from a small group that provides the broadest scope of information. In selecting the research group decisions have to be made about four key factors:

- the size of the research group. Deciding on group size requires a balancing act between keeping the group manageable for the chosen research style and the researcher's capabilities, yet large enough to be regarded typical for transferability purposes
- the scope of information sought and parameters of the group. Credible and appropriate selection of the group relies on the researcher clearly and accurately defining the parameters of the wider population beforehand
- access to the group. Access has to be practicable and permission sought. In some sensitive areas, for example, sexual abuse victims or young peoples' views on contentious societal issues such as evolution versus creationism, access may be problematic and careful negotiation over the nature and extent of access, and the release of findings, may be necessary, and

- the selection strategy to be used. Selection strategies are often purposive in interpretive studies, where the researcher deliberately selects a particular group within the larger population (Ball, 1984), and thereby avoids representing the whole group or randomness (Guba & Lincoln, 1989). Selection criteria tend to emerge as the study evolves and choices are ongoing, serial and contingent on previous findings. The wish is to present a faithful picture of a natural setting instance from which the reader makes comparisons to his or her own context.

3.2.5 Triangulation

Triangulation is the term used to describe the situation when two or more methods of data collection, or sources of data, or several different observers or investigators are used in a research project (Cohen et al., 2000; Coll, 2000: Keeves, 1998), to support or contradict an interpretation (Pitman & Maxwell, 1992). The use of triangulation has the potential in a naturalistic study to increase the trustworthiness of the investigation by providing rich detail about the complexities of human behaviour through data gathered from more than one standpoint. Both qualitative and quantitative data can be utilised. The power of triangulation as a research tool lies in its ability to reduce data, researcher, or method subjectivity that may be present (Coll, 2000; Erickson, 1998). The method allows the crosschecking of data from different sources and the assessment of the credibility of individual accounts resulting in a convergence upon interpretation of the human events being investigated (Bell, 1999; Keeves, 1998). Triangulation in interpretive studies commonly yields different types of data, but if there is a high degree of convergence or correlation between the outcomes of the different methods, the researchers can have greater confidence in the data's ability to give a credible picture of the situation (Erikson, 1998; Yin, 2003). This is particularly so "when the methods of data collection are at their most disparate" (Coll, 2002, p. 3). On the other hand inconsistent and contradictory outcomes of triangulation should not be discounted because they do arise out of different perspectives of the same event. They may in fact ultimately provide deeper understanding of the issues being studied (Coll, 2000).

3.3 Research Design

Once research goals have been operationalised into a series of issues or questions that can be investigated in concrete terms decisions need to be made, not only about the theoretical framework that guides the inquiry and methods, but also about the research styles, approaches, and instruments for collecting data most appropriate for the particular study.

Ball (1984) comments, from experience as an emerging researcher carrying out an in depth naturalistic study of a large school, that the final choice of methods cannot occur until the researcher is actually involved in active fieldwork. Ball found that his initial exploratory work opened up new research possibilities, and that the learning and perfecting of specific skills, such as interviewing techniques, and the necessary trialing and modifying of chosen methods all served to alter the research design as his naturalistic investigation proceeded. He advised that a naturalistic research design must also have the flexibility to allow methods to be devised on site in response to the demands and complexities of the people and organisations being studied.

As I neared the stage where final planning decisions needed to be made, I came to appreciate that my research design had to be by its very nature flexible and responsive to change, to ensure the trustworthiness and authenticity of the findings. I was also aware that the resources available to me for this study were limited, especially time, personnel, and cost. Working full-time placed restrictions on the time I could participate in observational research, and the small ‘window of opportunity’ available for me to engage in fieldwork and gather data would considerably influence the design of the study and the methods used (Keeves, 1998). The requirement in observational work for prolonged data gathering periods to assure dependability presented me with some difficulties, so to meet this trustworthiness criterium I explored other options for data collection. The use of questionnaires was one possibility, but the designing and trialing work necessary to maximise the validity of survey tools was considerable (Coll, 2002). Instead, I sought a compromise situation involving a short period of intense observation that was manageable within my time frame, yet yielded sufficient data to generate dependable findings. After due consideration of the best option for meeting as many as possible of the criteria of a trustworthy interpretive study, within my resource constraints, I made the pragmatic decision to use a carefully bounded case study approach (Adelman et al., 1980). This approach as described in the next section, would allow me to concentrate on a specific situation, to employ a number of different instruments gathering data from a range of sources, and hopefully to reveal the various interactive processes at work (Bell, 1999).

3.3.1 A case study approach

Review of a selection of recent interpretivist studies in classroom settings, both in New Zealand and overseas, shows quite a strong preference by these researchers for a case study approach to the overall project (Ball, 1984; Boaler, 2000; Carr, 2001; Cowie & Bell, 1999; Haigh, 2000). This style of educational inquiry is reported to provide researchers a more complete picture of educational practice than more conventional modes (Atkin & Black, 2003), especially since it requires the study to be done in context. A case study is "the study of an instance in action" (Adelman et al., 1980, p. 49) that provides a unique example of real people in real situations, and is typically focused on a system or unit with clear boundaries (Burns, 1994). In educational studies this could be an individual student, a group of students, a class, department or whole school. The emphasis is holistic, where coming to understand the context in which educational events of interest occur enables them to be understood more deeply (Atkin & Black, 2003; Keeves, 1998).

The case study approach is frequently used in order to portray some wider, more general principle. In justifying case studies as a means of carrying out research Burns (1994) indicates that focusing on the "idiosyncratic complexity" (p. 313) of a particular unit under study, by probing deeply and analysing intensively the many phenomena that make up the unit, case studies can generate evidence that enable generalisations to a wider population to be formed by the reader. Anecdotal evidence may serve to illustrate these general findings. To enhance comparability Yin (2003) says the presentation of a multiple-case study, even a two-case study, can increase the chances of producing more robust results than a single case, therefore enhancing the power of analytic conclusions – a form of replication. Cases studies are also the preferred strategy when attempting to find answers to *how*, *what* and *why* questions (Burns, 1994), particularly in evaluative research dealing with the processes and outcomes of programme implementation (Yin, 2003). Haigh (2000) also identifies the case study as a suitable strategy in educational research for enlightening educational policy and informing practice.

In observational case studies of an operational curriculum, a variety of data collection methods and sources can be used to intensively and directly observe classroom life (Burns, 1994; Keeves, 1998), although qualitative investigative methods seemed to be most favoured (Atkin & Black, 2003). For example, in case studies investigating formative assessment in science classes, Cowie and Bell (1999) used participant observations of teachers and students, audiotaped student interviews and examination of documentary data such as text and teacher planning. Participant-observers are defined as observers who have no pre-conceived ideas of what precisely they want to observe – they observe happenings and behaviours as they occur in the situation under study and record these observations as soon as possible (Bell, 1999). To illustrate, in the Cowie and Bell (1999) study, field notes including head notes (researcher's reflective thoughts), were a means by which observations were recorded for later analysis to augment the audiotapes. Haigh's (2000) study of investigative practical work in Year-12 biology programmes used similar methods, extending to the examination of written materials to departmental meeting minutes, Education Review Office (ERO) assurance audits reports, and the results of a teacher survey. Haigh used an initial questionnaire to gather information on the inclusion of problem solving in teaching programmes from over 250 teachers in the wider school region. These results were used as a basis for designing a second questionnaire that was used with teachers in the case study. Linsell (1999), who was investigating students' thinking in a constructivist mathematics classroom, combined video recording with stimulated recall interviews. Students were questioned in interviews about their thinking in class, whilst watching a video recording of the particular lesson, to gain a fuller understanding of the mental constructions they were making.

These examples of educational studies from the literature indicate the range of potential data collecting tools and sources that could be considered and utilised within a case study approach. To gather the wealth of quality data needed to provide the necessary thick and detailed descriptions of the setting under study, and keep the project manageable, I needed to make a careful selection for my research design from this range. As the project progressed other tools and sources were considered and incorporated where needed. The next section describes the research design decisions that I made, along with the reasons for their choice.

3.3.2 Data collection methods

After careful appraisal of factors related to the trustworthiness of the study, and the realities of time and logistics, I made the decision to do two case studies. My sampling was to be purposive (Cohen et al., 2000), primarily to keep the project manageable. Making further sampling decisions, to define the case or unit of analysis (Burns, 1994), required careful consideration of a potentially extensive list of parameters because of the heterogeneity of New Zealand schools. Parameters that immediately came to mind included school types, locality, year levels, participant roles, age, gender, ethnicity, years of teaching experience and science departmental size. I narrowed the field by considering parameters that were being highlighted in research and in national and international monitoring programmes of student achievement as contributing factors to differences in student achievement (Alton-Lee & Pratt, 2003; Carr et al., 2003; Harlen & Crick, 2003; MoE, 2002b; 2002c; 2004b; Pratt, 199;). Specific parameters were chosen in recognition of the low achievement of Māori compared to non-Māori; gender differences in student achievement and in rates of participation; the quality of teaching; learning for qualifications, home language and location. Although results from the Third International Mathematics and Science Study (TIMSS) [MoE, 2002c] showed rural-urban locality did not appear to affect student achievement in science there were indications in that study, and from the Programme in International Student Assessment (PISA 2000; MoE, 2002b) and the National Education Monitoring Project in Science (Crooks & Flockton, 2004), that decile rating was a related factor. The decile rating of a school is based on a socio-economic index that takes into account household income levels, categories of employment and the ethnic mix of a school. Generally a school is likely to have a wealthier and more monocultural school community with a high decile rating, on a 1-10 scale, than a low decile school. I therefore included decile rating, together with location, as a parameter to consider in choosing schools.

My sample selection would necessitate choosing groups for a specific purpose, that is, on the basis of their typicality (Cohen et al., 2000). This process first involved locating schools where the particular parameters chosen for this resulted in different school populations and educational contexts, that were typical of the kinds of school that do occur in New Zealand. Once a school was selected for the study then the second step involved choosing a group of students within that school who could best meet my selection criteria.

In each case study I therefore planned to involve a student group of about 4-5 students from a Year 11 Science class studying for NCEA from separate high schools. The two high schools were to include:

1. a high decile co-educational urban high school with a significant Māori population, and
2. a low-medium decile co-educational urban high school with a significant Māori population.

As a former science teacher and facilitator in the region, and as an Education Reviewer with contacts at the local university and the Ministry of Education, I was knowledgeable about most secondary schools and their principals and heads of science departments in the local area. I was also aware of the expertise and experience of many science teachers in these schools. This knowledge assisted me in selecting schools, and teachers, to match my research parameters.

To avoid potential tensions and/or misunderstandings that may arise as a result of inviting participant involvement it was important for me as a researcher to consider, identify and seek the support of all potential stakeholders in this research before participant invitations were issued. Apart from the participants themselves, stakeholders included fellow teachers and students at the school, parents/caregivers, school management and board trustees. To secure their understanding and/or permission for the project I approached these identified groups using a carefully planned sequence and in an ethically appropriate manner (see attached letters in

Appendices C, D and E). Once these stakeholders sanctioned access to the participants I was able to start the process of inviting participant involvement. In this process I again used introductory letters (see attached letters in Appendices F and G), but also included face-to-face meetings with potential teachers and student groups. Teacher involvement was secured first, and both teachers assisted me in selecting students to invite to participate.

Two high schools meeting my criteria were selected and invited to participate, but one school declined the invitation. A replacement school was found, but not all the criteria for the first choice could be met. The school was large, urban with a significant Māori population, but was single-sex boys with a decile rating of 8 instead of 5.

In teacher selection, factors such as years of experience, knowledge and philosophy of teaching were parameters to consider, but in reality the key deciders proved to be recommendations from the school management (heads of department) and the willingness of the teachers to participate in the study. I had hoped to select from a range of recently qualified teachers to those with many years' experience, with varied teaching styles and of different gender. In the final selection I was fortunate to secure as participants two female teachers with similar teaching styles, one in her third year of teaching and the other with eight years of experience. While the similarities between the teachers reduced the variability in the parameters defining each case study, it did offer scope for possible parallels and differences between the cases studies to be identified and explored.

Choice of student participation appeared relatively straightforward, since students were to work in groups for the topic study, and the students in groups whom the teacher thought suitable for my research purposes were willing to participate. The composition of these groups met most of my criteria, except for the inclusion of Māori students. Although there were Maori students in the classes of both cases studies, they were few in number and worked together. I made the decision at the time that asking one or two Māori students to break links with their usual groups and work with other students was not desirable since this did not reflect their normal working conditions. However, a fact I did not fully appreciate at the time was that the final student groups, chosen as participants in both cases studies, did have some members who had not worked together before.

The choice of Year-11 NCEA science classes would give me an opportunity to view teaching and learning within the context of a new qualification structure, and provide a source of quantitative data through achievement standard results. The rationale for collecting data, during the teaching of Science Achievement Standard 1.1 *Carrying out a practical investigation with direction* in each case study, was to facilitate comparability across the case studies in terms of the intended science learning, and so increase the dependability of data. This decision would also help me to keep the boundaries of the case studies well defined (Burns 1994), allowing me to explore the nature of students' learning during and about practical investigatory science. The time frame was set, yet sufficient to gather enough rich and detailed data for analysis without being overwhelming.

The case study approach enabled the triangulation of data collection using different methods, instruments, and sources. Data collection in both case studies took the form of field notes and head notes, and audiotaping and transcribing of interviews and classroom activities. Sources of data thus included participant observations, participant interviews, document analyses and qualification results. In my professional occupation as an evaluator I had become very experienced at participant observation and interviews, and was aware of potential dangers such as observer bias, pre-conceived ideas and prejudice (Bell, 1999). In observing a small group I also needed to be aware that my presence could alter the students' behaviour (Burns, 1994). However, this was an area of my investigation for which I felt reasonably well prepared, and confident that I had the prerequisite research skills and abilities necessary to gather the largely interpretive data required for a thick and authentic description of the situations under study. I was able for the large part to participate in the groups' activities such that the participants relaxed, and for large blocks of time ignored my presence and appeared engaged naturally in their classroom life. This allowed me to also take on a non-participant role (Burns, 1994), and gave me the opportunity to take extensive field notes. At other times I conversed casually with participants, seeking some clarification while at the same time trying not to interrupt the natural flow of the lesson. There were occasions, however, in both cases studies where my presence did modify students' behaviour, and this was noted in the findings (see Chapter 4).

3.3.2.1 The interviews

Interviews are important sources of information in case studies, and have the potential to reveal important insights and other sources of data through the eyes of participants (Burns, 1994). Teachers and students were the research participants in case studies where the students were undergoing study for Science Achievement Standard 1.1 *Carrying out a practical investigation with direction*. Research for the case studies included blocks of observational time - a three-week block for Case Study A (12 one hour lessons), and a two-week block (8 one hour lessons) for Case Study B.

Interviews with participants occurred before, during and after the observational blocks in each case study, and where possible I engaged in short discussions with the teachers after a lesson. In both case studies post-class conversations proved to be infrequent, due to the teachers' busy timetables and my requirement to return to my employment. Teachers were interviewed individually, but student interviews were conducted with a group of students for each case study, rather than as individuals.

Interviews with both teachers and students were initially planned as a set of standardised, open-ended questions carried out in a semi-formal fashion (Bell, 1999) – a form of verbal questionnaire, with prepared prompts should they be required (see Appendix C). The interview was given some structure to facilitate data analysis later in the project, but was open enough to allow participants to become more like informants than respondents (Burns, 1994). It also gave me latitude to explain or rephrase the question should the interviewees not fully comprehend the question being asked, or take a different meaning to that intended in my question (Driver et al., 1996). In reality, these interview schedules tended to serve more as checklists to ensure all relevant topics were covered, because the questions were modified according to the responses individual interviewees gave in the context of each case study. Where possible these interviews were conducted in quiet environments (Burns, 1994), but on a few occasions the circumstances of the interview meant that interruptions could not be prevented. The interviews were audiotaped, and transcribed by experienced, competent confidential transcribers. I then checked the transcripts against the audiotapes, correcting any typing errors and supplying parts of participants' conversation that the transcribers were unable to decipher. Some words/phrases were barely audible but I was able to supply them because of my presence when the data was recorded and the context. On other occasions the terms used were technical and not comprehended by the transcribers, or misspelt. Copies of the transcripts were given to participants for verification, and for further comment/clarification if wanted or needed either by the participants or myself.

3.3.2.2 Field notes

During observed classroom lessons in each case study, I took extensive field notes in a designated notebook, focusing primarily on the study group's activities and those of individuals within the group. I also took notes whenever possible on other key participants within the class, such as the teacher, and on events or aspects of the classroom environment that may impact on the learning of my group, such as wall displays, level of classroom noise or interruptions. Many of these notes were verbatim comments or descriptions of actions, while some included interpretations in terms of the research questions and criteria of my study that occurred to me as I was observing classroom activity (Spindler & Spindler, 1992). During data collection periods I tried whenever possible to analyse these field notes by highlighting and coding, but this proved difficult to do because of the large volume of data coming in over a short period of time.

In a separate notebook, I kept a diary in which was recorded my reflective thoughts or head notes (Cowie & Bell, 1999), and time spent on various aspects of my research. The head notes contained my developing ideas about emerging theory, possible methods of data analysis, and further issues to explore or questions to ask.

3.3.2.3 Audiotaping of teaching and learning sessions

Classroom lessons in each case study were also audiotaped. A microphone was placed as close to the group under study as possible without being too obtrusive, and left running for the whole lesson. The teachers' voices were generally quite audible and most of their verbal comments were caught on tape. However, the students' voices were frequently inaudible, especially during practical work when the overall noise in the classroom drowned out some individual comments. Also many of the students tended to speak quickly, or to mumble in these situations, and it was difficult for the transcriber to differentiate one student voice from another. At times, the microphone proved to be a distraction for some students, while for others it appeared to reduce the spontaneity of their classroom conversations. In each case study at least one or two lessons failed to be recorded due to technical difficulties, or operator mistakes meaning that these particular audiotapes were of little value.

Despite the difficulties and limitations of this form of data gathering, the audiotapes did add to the general pool of data collected in each case study, often corroborating other data or adding forms of data not evident from other sources. A few helpful quotable contributions were gained from these tapes.

3.3.2.4 Examination of relevant documents

A variety of documentary material related to the teaching and learning occurring in each case study was examined to gather relevant evidence. This documentation included:

- the *Science in the New Zealand Curriculum* document
- Science Achievement Standard 1.1 *Carrying out a practical investigation with direction*
- school schemes of work, including guidelines for delivery and assessment of Science Achievement Standard 1.1
- student workbooks, notes and assessment pieces (formative and summative items), and
- textbooks used by teachers and students.

The next section details how the methods of data analysis evolved as data collection progressed, and the thinking behind this process.

3.3.3 Data analysis

In interpretive studies using qualitative methodologies data analysis is ongoing and “begins as one is negotiating entry to the research site” (Erickson, 1998, p. 1162).

Since the case studies generated large amounts of information, I planned to attempt analysis as early as possible to select out significant features for future focus (Cohen et al., 2000). With my research questions in mind and using a process of recursive review (Erickson, 1998) I intended to carry out ongoing reading, sifting, grouping and regrouping of this information into sets of data to begin to gain understanding of the salient features of the situation. Then, using the understanding gained though this coding and re-coding of data, I could begin to generate themes to fit the situations I was researching.

In establishing my sets of data for analysis I had turned to my research questions for guidance and to my understanding of thinking and learning processes gleaned from the literature. I decided to adopt the strategy of seeking answers to the *what, why* and *how* questions (Burns, 1994), by drawing out of my data instances that demonstrated *what* students were learning, *why* they were learning and *how* they were learning. For these three aspects of my research questions to serve as the basis for my first analytical units, the parameters for each unit needed to be defined so data could be readily distinguished and classified.

To achieve this differentiation of the *what*, *how* and *why* of learning, I turned to the literature for insights. The literature reveals that interpretive based studies increasingly place emphasis on the social and language aspects of classroom practice to gain understanding of learning and inquiry (Haigh, 2001). Nuthall (1997), for example, suggests the amalgamation of sociocultural and linguistic perspectives into a cognitive constructivist model of the development of thinking processes to enhance understanding of learning. Hence to further develop my framework for data analysis I explored how linguistic and social processes within the classroom influence the ways students construct and retain knowledge. Students were viewed as intentional participants in classroom activities (Cowie & Bell, 1999) and the analysis was to concentrate on their perspectives of classroom reality. In this segment of my investigation I decided to survey constructivist, sociocultural and linguistic theories of teaching and learning (Nuthall, 1997) to find suitable parameters for my analytic units, since these theories have in common the view that knowledge has to be personally experienced (see Section 3.1.2 on naturalistic and interpretive paradigms).

3.3.3.1 The 'what' of student learning

In the *constructivist* view of learning students experience changes in what Leach and Scott (2003, p. 92) term the "mental structures" of individuals, that is, their concepts, schema or mental models. Individual learners construct their own knowledge motivated by the need to make sense of experience in light of their existing understandings. In constructivist terms, *what* a student learns during particular teaching and learning episodes are those science concepts, skills and understandings he/she has actively personally constructed as a result of the classroom experiences (McMillan, 1995; Skamp, 2004). The teacher's role is to provide learning experiences that enable students to construct knowledge as close to accepted science knowledge as possible.

The *sociocultural* stance on learning is that "thinking and learning are not seen as an activity of the mind in isolation, but rather as part of, or constituted by, the visible social interaction that takes place between members of a community" (Nuthall, 1997,

p. 701). What counts as knowledge is situated in the practice of that particular community and defined in social interactions (Barnett & Hodson, 2001; Black, 2001). Scientific concepts are cultural products that have been validated through rigorous and complex empirical investigation and social processes performed by members of the scientific community (Leach & Scott, 2003). Individuals could rarely discover or perceive such concepts without social interactions. Even the reading and interpreting of text (also considered cultural products) requires an individual learner to function in a social context in order to learn. Thus from a sociocultural perspective, learning is a process of enculturation where an individual develops the capacity to interact with other members of the community and participate effectively in its activities. Through social and cultural processes students in science classrooms learn by co-constructing understanding with their more expert teachers (Haigh, 2001), and fellow students come to learn viable science concepts through social reinforcement (Leach & Scott, 2003). In this view *what* students are learning in science could be the concepts, skills and practices that an expert scientist possesses, if the student as novice works in partnership with the teacher as expert (Hodson, 1996; Nuthall, 1997; Tytler, 2003). This view of learning is closely linked to the concept of situated cognition (Hennessey, 1993; Brown et al., 1989) which recognises that ways of knowing differ from one community of practice to another, and that learning is a process of enculturation into the ways of thinking of members of that community.

The *linguistic* perspective acknowledges the acquisition of language as a semiotic process (one of making meaning) that is central to all learning. Language is the means by which concepts are introduced and discussed by learners on the social plane, and the tool for individual thinking once concepts are internalised. There is 'continuity between language and thought" (Leach & Scott, 2003, p. 99). Linguistically then, the *what* analytic category could include data which demonstrate the acquiring of the scientific social language and speech genres, that are the way of talking and thinking within a scientific community, and using them appropriately in various situations.

This survey enabled me to determine a set of parameters that would define my unit of analysis known as the *'what'* of learning and accommodate the constructivist, sociocultural and linguistic perspectives on learning. My working definition for the *'what'* of student learning included those scientific concepts, skills and procedural knowledge (Bell, 2005; Duggan & Gott, 2002; Hodson, 1995; Skamp, 2004) students were acquiring and demonstrating through their words and actions during teaching and learning episodes related to the Science Achievement Standard 1.1. *Carrying out a practical investigation with direction.*

As my analysis progressed, however, the data exposed aspects of the practices and ways of knowing of 'school science' that differed from 'real' or 'authentic science'. These findings caused me to broaden the *what* parameters to encompass understanding of the nature of both authentic science, (i.e., that practised by scientists), and school science (i.e., that practised by participants in the school setting, including techniques for successfully meeting science assessment requirements in the NCEA qualification). This broadening of the parameters was an example of deciding progressively during recursive review which information to attend to further (Erickson, 1998). There was also a growing realisation on my part that in the *what* category I was including illustrations of both the learning intended by the teachers (the implemented or enacted curriculum) and qualification assessors, and the actual learning achieved by students. I needed to differentiate between the two forms of intended learning because the data depicting intended learning also relates to another of my research questions, that is, the match between the intended learning (national policy and teacher implemented) and the operational curriculum for students. This need led to the creation of subsets of data (i.e., the learning depicted in the intended curricula of national policy, NCEA departmental guidelines and classroom teachers' planning, and the learning experienced by students) within this major analytical unit to facilitate the task of determining the match later on when interpreting the findings.

3.3.3.2 The 'why' of student learning

Instances of *why* students are learning are characterised by the circumstances that led to students achieving that learning. In constructivist terms I would be looking for the conditions that prompt a student to restructure his/her own previous knowledge and beliefs, such as tasks that motivate students to engage in trying to make sense of their experiences. Or it may be the circumstances that cause a student to cling on to pre-existing ideas, such as alternative conceptions. Sociocultural and linguistic considerations of why students learn to recognise how the involvement of other persons promotes or hinders learning. For example, interacting positively with peers to co-construct understanding is increasingly recognised as a powerful aid to learning (Bishop & Glynn, 1999).

When teachers interact with students, they commonly use a range of purposeful acts of teaching or instructional strategies (MoE, 2003c). These strategies can include modelling and demonstrating ways of knowing in science, and giving students the opportunity to use the cultural tools of science and engage in its practices. Helping students to activate prior knowledge, providing activities with appropriate kinds and levels of challenge and giving timely and appropriate feedback are means by which teachers can scaffold learning and enculturate students (MoE, 2003c; Torrance & Pryor, 2001). Relationships with peers, school culture, parental and societal expectations and personal interests are examples of other sociocultural factors that could emerge in the data as reasons why/why not students learn.

3.3.3.3 The 'how' of student learning

What learners do in order to learn was the key parameter I chose to define the *how* set of data. This approach to the *how* of learning focused on the thought processes and actions attributed to students as they learn (MoE, 2003c). Any act, or form of engagement in thinking, by students in a social context that appeared to contribute to their acquisition of science skills, knowledge and attitudes I therefore interpreted as an instance of *how* they learn. Examples of such instances included students imitating others' actions, memorising, questioning, listening, making connections with prior knowledge by recalling, practising, developing the ability to apply their learning in new contexts, and verbalising thoughts. Many of these behaviours and operations are generic to all learning, and involve the learner as an active participant.

3.3.3.4 Categories within categories

As I continued to scan the information sources looking for points of interest, more possibilities for sub-categories began to surface (Cohen et al., 2000). For example, absenteeism from lessons, failure to bring texts to class and off-task social chatter were all circumstances that I observed having the potential to adversely affect science learning. This sub-group I labelled as ‘barriers to learning’ under the broader heading of *why* students learn. Even within this sub-group, further distinctions became evident as I recognised some barriers to learning were linked to social interactions, while others were caused by limitations of the physical environment such as hot and stuffy classrooms or the layout of laboratory benches and chairs, and so on.

I also found later, when interpreting the findings, that it was very difficult at times to distinguish the *why* aspect of particular data from the *how,* and in fact unhelpful for the purpose of the study. To facilitate the interpretation and discussion of these findings (Erickson, 1998) I interwove the two aspects into one section in a fashion I believe maintained the integrity of the data and findings.

3.3.3.5 Participant validation

As part of on-going analysis and member checking, I had originally planned to supply the participants with draft interpretations. This planned practice did not occur during the data-gathering period because of time factors and my inability to do much data analysis until some time after the on-site work. However, I was able to feedback some analysed data to participants in the form of a questionnaire for students late in each case study (see Appendix I). I devised the questionnaire as a data-gathering tool to gain assurance from student participants that my initial categorising of findings was an accurate representation of their views, and to confirm some trends I could see emerging from classroom observations and interviews. In effect, this tool gave student participants the opportunity to validate some of my interpretations during the study, namely instructional and learning strategies that they were identifying as effective ways to learn. To some extent I was also able to convey some tentative interpretations during the later interviews by discussing points and raising questions arising from previous interviews.

The last section considers the ethical considerations that needed to be taken in account to ensure the safety and wellbeing of all participants in this study, while still maintaining the trustworthiness of the research.

3.4 Ethical Considerations

Today all researchers of human behaviour must be aware of and take into account the attendant moral and ethical issues that pervade their work (Cohen et al., 2000; Snook, 1998). Researchers have obligations to act in the best interests of all people involved in or affected by the investigations and make sure that no significant harm is caused (Erickson, 1998). This requires investigators to act ethically, which "is to act the way one acts towards people whom one respects" (Graue & Walsh, 1999, p. 55). It is about the attitude they bring into the field and to the interpretations they make.

Ethical concerns will often place researchers in dilemma situations. Typically in such situations a balance must be struck between the demands placed on them as professionals to pursue the truth, and the potential threat of their work on the rights and values of their research subjects (Cohen et al., 2000). Ethical thinking needs to combine reason with a sense of rightness since "a design which prevents the gaining of truth subverts the ethical legitimacy of the undertaking … for research to be ethical it must be valid" (Snook, 1998, p. 5).

Ethical issues can arise at any stage within an investigation. To deal successfully with them a researcher needs to be well prepared, and very aware in their own mind of the nature, direction and scope of their work before they begin (University of Waikato, 2001). Thinking beforehand about how the issues of privacy (McBride, 1994), storage of personal information, anonymity, confidentiality, conflict of interest, betrayal and deception can give rise to moral and legal dilemmas, and is good ethical research practice. Seeking informed consent, gaining access and acceptance, behaving honestly and truthfully, and avoiding harm to the respondents are all essential procedures in ethical research (Graue & Walsh, 1999; Snook, 1998). The University of Waikato places a high priority on ethical considerations in research, and before my data-gathering phase began I was required to submit a formal written application that outlined how key ethical and legal issues, that may or may not arise in the proposed research project, were to be addressed. Any human research undertaken at the University of Waikato must have prior approval from its Human Research Ethics Committee. The following sections describe some key ethical considerations in my study and the steps I took, either to prevent ethical issues arising, or to address any that might arise.

3.4.1 Confidentiality

Credible assurance of confidentiality on the part of the researcher is essential to ensure the trust of all involved in the research. The small number of schools in the city and the surrounding rural area increased the potential for the schools and participants in the study to be identified. These risks were managed by ensuring information was not provided that could allow identification of the respondents. Thus information such as participant names, and addresses, and the name and location of the school, in which they teach and learn, were not disclosed. The manner in which I reported my findings also served to maximize confidentiality. For example, the assigning of pseudonyms to participants and schools was one means by which identification was prevented. Participants were also informed of their need to maintain confidentiality.

3.4.2 Informed consent

This process involved obtaining the consent and cooperation of the participants (teachers and students), and of the personnel with responsibility in the schools in which the study takes place (board of trustees and/or principal and senior management team, and head of department). Students' informed consent was sought rather that their parents since they were of a mature age (15-16). Participants were provided with all relevant information about the research project, so they fully understood the nature of their involvement and the implications of that involvement. A letter was sent first to the principal seeking consent for access to the school, teacher and students (see Appendix D). I considered it particularly important that the students understand and appreciate their role in the research. This provision of information was initially achieved through face-to-face meetings with participants. After teacher meetings I met with the whole class to introduce myself, and explain and discuss the project. Letters to all class members and teachers followed up these early meetings, seeking informed consent (see Appendices E, F and G). Information about the project was also provided to parents/care-givers of students involved as a matter of courtesy (see Appendix H).

Potential benefits and risks to participants, by their involvement in the research, were explained to participants and they were free to choose to take part or not. They were also able to make inquiries about any concerns they had with the procedures and have them answered, and had the right withdraw at any stage in the research process without prejudice to them. In Case Study B two students did withdraw from the study partway through, citing a desire not to continue, and this departure was able to occur without incident. Fortunately two other students willingly took their place. From all participants in both cases studies I was able to obtain their written consents.

3.4.3 Conflict of interest

An ethical issue that I needed to address concerned my position at the time of the study, as an Education Review Officer, and possible conflict of interest situations I might find myself in. As an education reviewer I was legally required to make unbiased judgements about the quality of education being delivered in New Zealand schools, and a professional code of ethics governed my behaviour (ERO, 2002). There was undoubtedly the potential for me to be perceived as holding a position of power over participants. This was to be avoided at all costs if trusting and open relationships were to be built. At times I could see things that caused me discomfort as a reviewer (Graue & Walsh, 1999), but these feelings had to be set aside to ensure the integrity of the research and the confidentiality of the participants. The onus was on me as a researcher to protect all participants and ensure that none of my actions caused them harm (Cohen et al., 2000). To minimise the risks it was appropriate for me to select schools that I had not reviewed, and to undertake not to review the selected schools for a period of time after the research is completed. This I undertook to do in the letters of consent signed by participants.

3.5 Summary

Building on the literature review of Chapter 2, this chapter provided a rationale for the theoretical framework and a description of the methodology developed for this inquiry. The methodology, based in an interpretivist paradigm, comprised a multiple case study approach utilising the qualitative research methods of participant observation, semi-structured interviews and document analysis.

Chapter 4 reports the research findings organised around the research questions for each case study in turn.

CHAPTER FOUR

CHAPTER FOUR

FINDINGS FROM THE CASE STUDIES

Overview

The focus of this chapter is the presentation of findings from the two case studies. Using narrative style and format, holistic accounts of the two case studies are given as they unfold over time. Each case study begins with a description of the contextual setting and an introduction to the participants. The storylines are then each given a metaphorical framework to help the reader begin to distinguish key similarities and differences between the two educational situations. Both narratives weave the '*what*', '*why'* and '*how*' of student learning into the storylines, in a manner designed to facilitate discussion of the research questions in the following chapter.

4.1 Background to the Storylines

Two of my research questions are concerned with determining the content (the '*what*') of certain curricula, in particular, the intended curriculum of national policy portrayed in the SiNZC, the teacher's intended curriculum (implemented curriculum), and the student-experienced curriculum (operational or achieved curriculum). In my methodology I have established parameters for identifying and defining the curricular content on which my study will focus. These parameters initially limited the content to those scientific concepts, skills and procedural knowledge SiNZC was promoting, and teachers were delivering and students acquiring during teaching and learning episodes in the case studies related to the Science Achievement Standard 1.1. *Carrying out a practical investigation with direction.* As a result of findings emerging from the data, the procedural knowledge aspect was expanded in the enacted and student-experienced curricula to encompass aspects that could perhaps be considered more part of the hidden or implicit curriculum (Eisner, 1994), that is, being able to distinguish ‘school science’ from ‘authentic science’, and mastering techniques and strategies for meeting the assessment schedule requirements of tasks for this internal NCEA standard.

The story-line approach arose in response to my need as an interpretivist researcher to paint a picture that will allow the “voice of the research participant to be heard for others to reflect on” (Bishop, 1997, p. 30). I have also introduced a metaphorical element into each story to help the reader develop a sense of both the uniqueness of each case study and the similarities. Each storyline describes the educational context before introducing the participants and their perspectives about teaching and learning in general. The narratives then focus on the concepts, skills and approaches to scientific investigation that the teacher intends students to learn. The subsequent curriculum that the teacher actually delivers and the curriculum students come to learn are interwoven into the storylines. My interpretations of the nature of these curricula are based on what I observed in classroom activities and students’ written assessment results, and heard as students and teachers articulated their views during interviews. I have attempted to build a progressive picture of the students’ learning experiences over time in these narratives, selecting themes and aspects from the data that serve to highlight and illustrate the nature of their experienced curriculum.

4.2 The 'River Valley Boys' High School' Production

4.2.1 The school

River Valley Boys High School (a pseudonym) is a large, urban, single-sex boys' school. At the time of the study the school's decile rating was 8, indicating students were drawn from a predominantly upper range socio-economic community. The three largest ethnic groups were New Zealand European (77%), Māori (12%) and Pacific peoples (3%). The prospectus describes a school culture characterised by a strong belief in traditional values of academic excellence, self-discipline, respect, service and a competitive spirit. The active promotion of these values is noted in a recent Education Review Office (ERO) report, highlighting the school's strong emphasis on celebrating excellence in all areas of student endeavour and traditional, 'sound' classroom teaching.

At the time of this study the school was in its third year of implementation of the NCEA qualification at Level 1 (Year 11), and the 2003 ERO report indicated that implementation had gone reasonably well at the departmental level. However, it was reported that some teachers felt NCEA assessment was driving the curriculum, and teacher workload had increased due to lack of relevant resources, such as assessment items, for the qualification. The report also commented that students needed time and experience in the new qualification to gain understanding of its nature, with some students in particular experiencing frustration at the requirements needed to achieve at the 'excellence' level.

In the school, students from the Year 11 science cohort were effectively 'broad-banded' into three ability groups on the basis of their science option choice. One band comprised high-achieving students who studied an academic course designed for the International General Certificate of Secondary Education (IGSCE), an overseas qualification offered by the Cambridge International Examinations System. A second band comprised students in essentially vocational courses that used unit standards and/or achievement standards for qualification purposes. However, the majority of the Year 11 science cohort was in the third band, studying a course consisting of six science achievement standards (ASs), including the internally assessed AS1.1 and five externally assessed ASs. The Year 11 science class selected for this case study came from this third band, and was described by Jenny (a pseudonym), the classroom teacher, as a 'mixed ability' group. A range of ethnicities existed in the class but the mix differed from that of the whole school, in that the proportion of Asian in the class (33%) was greater than that of Māori (10%).

All science lessons in the study were taught in a large, purpose-built laboratory, with students sitting at high benches permanently fixed in rows parallel to the front of the room. For much of her teaching, the teacher was based at a bench on a platform at the front of the classroom, although she frequently moved around the classroom when students were doing individual or group activities. Behind her was a whiteboard and pull down screen with an overhead projector to the side. Large wall display areas were available on three walls and held a mix of commercial posters on various science topics, and publicity notices for upcoming science events such as a regional science fair. There also was one example of a student poster. Sets of common laboratory equipment and chemicals were stored in labelled trays or on shelves at the front of the room - all were accessible to students when needed. Sinks and gas taps were situated on student benches.

4.2.2 The teacher

The classroom teacher Jenny held a Master's degree in science majoring in genetics, and was in her eighth year of teaching. Before arriving at River Valley Boys' High two years earlier, she had spent her first five years of science teaching at a state (public) boys' school in another city. This teaching experience included curriculum leadership, taking responsibility for developing a differentiated programme for low ability students, and one year spent coordinating the Year 11 programme in the school. In her second year at River Valley Boys' High she was made teacher-in-charge of Junior Science, a position she held at the time of the case study.

Jenny had taught Year 11 science both prior to and since the introduction of NCEA, and was into her third year of teaching Year 11 science under this new qualification system. In the year leading up to Level 1 NCEA implementation, she had attended professional development courses provided nationally for all teachers by the New Zealand Qualifications Authority (NZQA). In subsequent years she attended similar national professional development courses for Levels 2 and 3 NCEA in both science, and specialist sciences of biology and chemistry.

In her initial interview Jenny clearly articulated her philosophy of teaching and learning in science, and the rationale for her views. An essential feature of her philosophy was "that every student should be achieving to the very best of their ability, no matter what their ability is". This she felt should occur no matter the ability of the student, so "if they are capable of 40%, they should be getting 40% and it is my job to ensure that they do that … what is the most important thing is that students achieve to their potential and even more so or higher than they ever thought". One consequence of her philosophy of teaching was that she exerted significant control of her classroom, meaning that she ran "a fairly quiet classroom. If I'm talking I expect it to be absolutely silence and I expect all students to respect each other, listen to each other".

These views appeared to be influenced principally by her personal experiences as a student during her own schooling: "I don't think I went to a very good school myself. My teachers didn't push me to achieve. I wasn't pushed to achieve at home. So I was pretty slack in terms of attendance and effort, homework, everything. And I certainly didn't achieve what I was capable of". Consequently in her tertiary level studies "I found, I didn't have a very good background in some of the areas that I'd been wanting to do and ultimately it probably changed my career because I then couldn't do what I wanted to do". She attributed the gaps in her own education to a lack of clear direction from her school (and family). The importance of the school in providing strong guidance for students was reinforced during Jenny's initial teaching experiences in schools where she witnessed how the ethos of a school and good teaching could have a positive impact on student performance. In these schools she "saw how the students were directed to succeed academically within the classroom no matter what their ability, I saw the difference and that probably has altered my philosophy".

Assessment also strongly influenced Jenny's approach to her teaching role. She credited assessment with perceptions of high status both in New Zealand's education scene because "at the end of the day it's an assessment-based education system" and in the wider New Zealand society. She felt "everyone has to have their performance measured. It happens when you leave school as well, and your profession. So I think the students need it too really … assessment is really important". She felt qualifications impacted significantly on her students' lives: "They need to pass NCEA, Level 1, Level 2, Level 3 … If they don't achieve that then they still have the stigma that they have failed". As a consequence, empowering students to achieve in qualifications featured prominently in her approach to classroom teaching: "I think in class we need to keep them up to that [success in qualifications] a lot. I often start my lessons with a 10-question quiz type thing, so just to go over what we've been learning in class. So that constant repetition hopefully drums it in, into all of them".

When asked about her teaching style Jenny saw it as "teacher orientated", largely because "the availability to textbooks isn't that great, so I don't often set my students work to work out of a textbook". Textbooks were used chiefly for homework tasks such as problems. Practical work played an important part in her pedagogy, because she believed strongly that "hands-on" activity helped learning: "I'm very practical orientated, I suppose as a science teacher. Most of my classes will have a practical everyday if at all possible. I think that's what helps them learn in science hands-on … certainly not by me standing up in front of the classroom and talking". When asked if she viewed practical work as part of investigative science she responded affirmatively, citing instances where she gave students the opportunity to carry out practicals without initial teacher direction: "Often with classes I won't even introduce something. They'll just go ahead and do the practical, find out what's happening. I'll go around and interact with them quite a bit and sort of give them leading questions the whole way through, then we'll come back and discuss it". She considered these practicals were investigations, in the sense that she took a more "hands-off" approach. Jenny particularly favoured this approach with higher ability students, whom she believed could cope better because "they really need to know what they're looking for, I find, if they're going to achieve or realise what it is, when it happens".

In summary, Jenny had a very clear vision of her role as a science teacher. That role included a responsibility for ensuring all students achieve to their maximum capability, particularly in qualifications. This philosophy she sought to manifest in her teaching approach through a teacher-directed learning programme, focused on carefully structured and sequenced lessons, plentiful opportunities for hands–on practical work and regular monitoring of learning through assessment. Thorough preparation for qualifications also involved her students in regular practice of the specific skills and techniques needed to accomplish the assessment tasks used to decide qualifications.

4.2.3 The students

The students who accepted my invitation to be the study group were four New Zealand European males, approximately 15-16 years old. The data-gathering period was early in the school year, and Jenny recommended these four students to me as research participants based on her initial impressions of their actions and behaviours in class. Martyn, Peter, Mitchell and Eddie (pseudonyms) were students whom Jenny felt would work well together, and also be prepared to talk openly about their experiences. Later in the study period, another student, Sam (a pseudonym), joined the study group because he frequently interacted with group members in class, and was genuinely interested in being involved in the research.

Prior to this unit of work the students were not well known to one another, and had not worked together as a group. Of the four original members of the group, Martyn and Peter were the most forthcoming in their comments, and in their willingness to share their thoughts. They both proved to be active group members in the investigative work in class. Mitchell, on the other hand, tended to be reticent in the interviews, but willingly conveyed his views in a written questionnaire I developed for the last interview and involved himself in the group work. Eddie proved to be a reluctant participant in interviews and was rather withdrawn, offering few insights into his thinking. In class his involvement in the group's investigative work varied – on some occasions he displayed disinterest, while on others he engaged in off task practical activities. However, during the formal assessment activities, especially the summative event, there were occasions when he became actively involved, even at times taking on a purposeful leadership role. Sam, the eventual fifth participant, was prominent in the whole class forum as a student who often engaged in discussion, frequently contributing ideas and asking questions. He was eager to accept my invitation to join in on the last interview.

In my first interview with the boys, most expressed general enjoyment of science especially the practical aspects. Martyn viewed science as "probably one of my favourite classes and I like the practicals", while Peter admitted he used to hate science and going to the classroom, "but now that I start to understand it more it's more fun". Eddie enjoyed "the practical part more than anything" and the others agreed, especially if explosions were involved. Mitchell's comment that "it's fun, blowing up stuff" was typical of the groups' views on the kind of practical work they liked. The group perceived topics like "rocks" to be not so interesting and consequently Martyn "never learnt about those and skipped that section".

When asked if they learn well in science Martyn responded affirmatively. He saw himself as able in science: "You just know instinctively … I've always been good at science, I just do better in science … I always do quite well in tests". Peter believed that he had become better at science: "When I started to listen more … taking more notes and then I started getting higher scores". He felt that "when you're getting it, you know what the teacher is on about … it just comes naturally". Mitchell commented though that he found science "quite hard … confusing sometimes", and he knew when he was not learning well: "I find that if you miss something that makes things hard. You're off it. You're not going to get it so it gets worse". However, Mitchell admitted that he normally did better in tests than he anticipated: "Like I think it's harder but I go in there and I do pretty good". Martyn recalled similar experience: "I found that. I think it's harder but then I go in there and I actually do quite well". Unlike his peers, Eddie was adamant that he did not know if he was learning well in science: "I don't know", and he did not elaborate on this despite further prompting during the interview. Martyn made a final comment about his surprising experience last year where he admitted that he "messed around all year, didn't write down any notes and then when it came to exams I still did real well! I still picked it up, most of it".

The interview then turned to discussion about their teacher Jenny and the approach she used in her teaching. The group's impression was that she took a similar approach for most topics: "Starting right at the basics of it all, and then it's building onto that … we usually write notes up for it then we carry on, maybe a practical. She will write it up there and explain it … how to do it and what to do". They described how Jenny gave quick 10 question tests about twice a week, and often made them try a few quick questions in their books "for each thing we've learnt about".

When I asked what helped them to learn, they agreed with one another that anything that involved going over work again helped learning. They cited strategies like diagrams, practicals, and models as helpful because they were "quite often another form of explaining it … different kinds of examples". On the other hand going over things too fast, when the teacher "just writes up the notes and then goes on to the next thing" hindered their learning, because she was "not explaining it". Sitting with friends could both hinder and aid learning in their opinion: "If you sit with your mates you tend to talk and you're not listening". But "sometimes sitting next to them is easier too because you can then talk about it and understand it, and they understand it". The students were a little divided in their opinions about the kind of teacher-student relationship that best promoted learning. Martyn felt a teacher "could be fun and tell jokes", and still be a good teacher – he didn't appreciate teachers who disciplined you for asking your neighbouring students questions in class. Peter didn't really think his relationship with the teacher mattered when it came to his learning, although he felt that "when they're nice to you it's easier". All agreed that a teacher needed to be knowledgeable about their subject. Mitchell thought a teacher "who knows more about the thing than they are actually teaching" helps him to learn because it's "easier to ask questions" of such a teacher. Martyn believed teachers who were aware that not all students necessarily had the same level of understanding in certain lessons, and who "explain it all and make sure people know", helped his learning.

Summaries, in the form of lists, mind maps or simplified notes, the students felt helped them learn particularly when revising before assessments like tests and exams. Mitchell recalled that in "Year 9 we had our own before every test … a whole page, and then at the exams you could just rip through those. You don't have to read everything in your book". Peter preferred lists because, unlike mind maps where "you've got like to look for it [in your books]", the information was all there: "You can just read through it". Martyn valued having a set of key points to remind him of what he had missed in the past. They all felt revision helped their learning. Martyn had good intentions but seldom realised them: "I might think I'll read a bit on the bus, then on the bus I never will". Paul felt revision went well if the teacher did it with them because "you can ask more questions, like if you don't understand it". Martyn agreed, provided the teacher directed the session by "putting work up and asking questions" to see if they understood.

The group had mixed feelings about family being able to help with learning. Peter said if he wanted help they gave it but this was not often, and Mitchell did get help but for mathematics, not science. Martyn's father knew a lot about science but Martyn had discovered that "I'll ask him a simple question and then all of a sudden he talks real complex and he gets carried away. One time I asked him about the formula of something and I ended up understanding how black holes work!" Peter did make a comment in a later interview that his parents helped him learn because they made him "do his homework".

The last topic discussed in the first interview was the NCEA qualification. Martyn and Peter confided that they were not overly confident about their ability to achieve learning success in NCEA. Martyn expressed concern because he had heard NCEA "makes smarter people look dumber and dumber people look average". He explained that he often got hard questions right, but missed easier ones, so he felt he was "going to start failing all NCEA". Peter, concerned about his chances of success in NCEA, cited his mathematics teacher who had said NCEA did not differentiate between students who had scored highly within a grade and people who had just scraped in. Mitchell appeared unworried about his learning for NCEA, and Eddie did not express any opinions.

4.2.4 The story as it unfolds

I have likened aspects of this story, as Jenny prepared her students for their summative assessment experience for Science Achievement Standard 1.1. *Carrying out a practical investigation with direction*, to an experienced director preparing her amateur actors for a one-night performance of a play. The intention is that this metaphor will convey to the reader Jenny's clear vision and focus on maximising her students' achievement on 'performance night', and her students' growing appreciation of what constitutes a good performance and their desire and anxiety to do well 'on the night'.

4.2.4.1 The director's interpretations and tactics

Interviews with Jenny, and observation of her teaching materials and actions in class, reveal that she drew on five sources to guide her teaching for Science Achievement Standard 1.1 *Carrying out a practical investigation with direction*:

- a section from a student workbook publication (Cooper, Hume & Abbott, 2002), featuring material focused on Science Achievement Standard 1.1 *Carrying out a practical investigation with direction*
- the school's departmental guidelines for the internal assessment of this standard
- her own research experience gaining tertiary qualifications in science
- her previous teaching experience, and
- feedback from her classroom teaching, especially information she obtained from the formative assessment activity including students' written scripts.

As might be expected, from the interviews with Jenny, the examination of her teaching materials and observation of her teaching, I found the specifications of the AS had a strong and direct influence on the content of the curriculum she delivered to students. This influence was brought to bear on her delivered curriculum via two access channels.

The first channel was created by Jenny's use of a section in the student workbook publication (Cooper, Hume & Abbott, 2002) as the basis for her teaching and learning sequence, and the source of most learning activities. The workbook was divided into seven sections, each section corresponding to a particular Level One Science AS. In their interpretation of the requirements of Science AS 1.1 the authors of the workbook had adhered closely to the standard's achievement criteria and explanatory notes, and exemplar materials provided by NZQA such as assessment activities, schedules and templates. The authors presented text and activities in a sequence designed to scaffold the learning of concepts, skills and procedural knowledge, which in their interpretation met the requirements of this internal standard. In classroom lessons, and the setting of homework tasks, Jenny followed this sequence and students covered most of the material and exercises.

The second channel involved Jenny's use of guidelines consisting of internal assessment activities and schedules developed by her science department. These guidelines were based on NZQA exemplars, feedback from NZQA moderators, and the collective experience gained by departmental members during assessment of the standard in preceding years. Teachers were allocated about 10 lessons to deliver and assess the standard. This time block approximated to two periods for teaching of content, four periods for the formative assessment exercise and four for the summative assessment event. As required by all departmental members for moderation purposes, Jenny took great care to follow the guidelines closely in her classroom delivery. She demonstrated this adherence particularly in the classroom sessions leading up to and following the formative assessment (a practice or mock assessment), and in the intervening period before the summative assessment. In her teaching approach Jenny focused on familiarising students with the background science concepts to the practical provided in the guidelines, and on the achievement, merit and excellence requirements outlined in assessment schedules. In class she frequently alluded to the generic requirements of the standard in different teaching and learning contexts, and directly referred to specific requirements when giving feedback and feedforward to the class after the formative assessment. In the lead-up to the summative assessment Jenny again carefully schooled the students in the requirements of the standard, by re-emphasising the generic skills and procedural knowledge needed for each step of the assessment task, and teaching the relevant background science concepts (to do with pendulums) as indicated in the departmental guidelines.

However, one aspect of the procedural knowledge in Jenny's intended and delivered curricula did not stem directly from the achievement standard's requirements, but rather from her personal philosophy about the nature of science and the nature of science pedagogy required in schools. In describing her approach to teaching scientific investigation she articulated a clearly developed notion of "good science", which she equated with science performed by "real scientists" in the world outside the classroom. She frequently referred to good science during her teaching episodes in the classroom. For example, during an episode of shared reading from the student workbook, as Jenny guides her class through the requirements of the science achievement standard, she asked: "If I get you guys to do an experiment I expect you to do it only once. Is that 'good science'?" When a student replied that he believed it was 'good science', she responded by saying it's actually "really bad science" because if the students were scientists "out doing an experiment yourself you have to do repetitions … to get any sign back your results are reliable". She explained that she is "actually teaching very bad science", but that time and resources constraints limit the experimentation they can do in class, so specially selected experiments known to work are used in schools "so you should get a good result only by doing it once". In this instance, Jenny was making the point that "real scientists" perform their laboratory investigations in a more rigorous manner than students can usually achieve in school laboratories because they are less constrained.

In interviews Jenny revealed the tension she felt between the reality of classroom teaching and the ideal she wanted to portray: “I always feel pushed for time so I tend to just usually give them the outline of the practical, this is what you do, whereas that’s not good science, they really should be planning what to do themselves and that doesn’t happen”. To resolve this tension she took a pragmatic stance, explaining that in her view a classroom practical has to be, out of necessity, different to real science. She justified this stance by explaining that classroom practicals serve a different purpose to those carried out in “real science”, that is, a pedagogical purpose rather than a scientific one. She maintained that in most classroom practicals “you’re not doing an experiment in the real world”, pointing out these practicals are not original investigations, but experiments, which have been carried out by many different people in the past in order to learn particular scientific concepts and skills. Jenny felt her students needed to experience these same experiments because they too “need to know the stuff which three million people have learnt before”.

Interestingly, Jenny’s felt Achievement Standard 1.1 *Carrying out a practical investigation with direction* did provide scope for students to practise good science in their practical work. When asked about what she saw as learning successes for the students taking the standard she commented that it gave students the opportunity to engage in real science, which they do not usually have in conventional classroom practicals: “To be honest I only really get them to plan and experiment when we do the planning experiments [for the standard] you know … it’s a different part of science, I mean you're trying to normally get them to get a particular result so they need to do that particular experiment, whereas this is teaching them in a real world". She felt the approach in the standard towards planning had merit because "this does teach them better to be aware of the variables in terms of what to control, more, far more than what we do in class … they really should be planning what to do themselves and, the fact that they have to undertake it an experiment themselves and a good experiment too". She valued students developing the ability to develop write clear plans "that anyone can come along and use and follow". Students being required to interpret their results in relation to the aim of the investigation she thought valuable because "that's what science is about … you've got to look at your results and then decide what your next experiment is going to be". Overall Jenny appeared to equate much of the planning, carrying out and reporting of the practical investigation that students did for Science Achievement Standard 1.1 with her notion of the 'good science' that scientists do: "I think they are learning some of the crucial skills of being a scientist better than they have probably in the past [under the previous School Certificate qualification]".

The key concepts, skills and procedural knowledge that Jenny intended her students to learn are summarised in Table 1. This content was identified in data collected from her interviews, observation of classroom lessons, teaching guidelines and notes, and the student workbook (Cooper, Hume & Abbott, 2002).

Table 1

The Teacher's Intended Learning (Case Study A)

Concepts	Skills	Procedural Knowledge
• 'Good science' (the science that real scientists do), and 'school science' (the portrayal or simulation of science experienced by students in school) • Fair tests • Purpose of an investigation as an aim, testable question, hypothesis or prediction • Variables - key, dependent and independent • Primary and secondary data, qualitative and quantitative data, reliability of data • Graph types (bar and line); graph components such as title, x (independent variable) and y (dependent variable) axes, units and values for axes, plotted points, and lines of best fit • Sources of error and systematic errors • Equipment names, types and purpose • Background/contextual science concepts to the investigation e.g., factors affecting rate of reaction.	• Designing, trialing, evaluating, modifying and carrying out a systematic plan for a fair test • Determining the purpose of a fair test investigation • Identifying, controlling, changing, observing and measuring variables • Choosing and using equipment appropriately • Determining appropriate range of values for variables • Repeating experiments • Recording and processing data – tabulating, averaging, graphing • Interpreting data, and recognising trends and patterns • Discussing findings, linking findings to existing science ideas and drawing conclusions in a written report • Evaluating the investigation in the written report (sources of error, improvements).	• Knowing how to plan a workable, fair test • Knowing that planning requires trialing, evaluating and modifying • Knowing why reliable data is needed and how to obtain consistent data • Knowing how to recognise and account for errors in measurement • Knowing that the findings should be linked to science ideas • Knowing how to work as a team • Knowing that the planning and carrying out of investigations required for Science A.S. 1.1 more closely resembles 'good science', than most 'school science' • Knowing how to interpret the template and assessment schedule requirements of tasks for the internal Science A.S. 1.1 at achievement, merit, and excellence levels.

4.2.4.2 Learning the lines and moves

My classroom observations in the first few lessons revealed only a few indications of the scientific thinking and/or learning that students might be achieving at this early stage of their course of study. The amount of data I could collect was limited because Jenny ran these lessons primarily as instructional sessions. The learning activities were conducted in a whole class, teacher-directed forum, but were interactive in the sense that students could respond to questions posed by Jenny and offer ideas, or ask questions themselves. Most of my student participants tended to focus on listening, reading and writing activities, although Peter and Sam did ask a number of questions. Jenny used the lessons to provide students with information about the teaching and assessment sequence for the unit and to teach them the key concepts, skills and techniques needed for addressing the assessment requirements of Science AS 1.1.

In the initial instructional session Jenny informed the students that the standard was going be taught in a block, including a formative assessment where they would be "learning how to do it", followed by a "revisit, then a final summative at the end of Term 1". She then introduced students to the standard through a whole-class, guided reading session of the first two pages of the student workbook (i.e., Cooper, Hume & Abbot, 2002). These pages provided the achievement criteria from the standard and explanatory notes elaborating the meaning of key terms in the criteria like 'purpose', 'workable plan' and 'sources of error'. Students took turns to read sections aloud and Jenny punctuated the reading with questions (mostly closed questions to do with recall and procedure) and further clarification, placing emphasis on the meaning of terms, aspects of experimental design, and what was needed to achieve particular grades. The following excerpts from a lesson illustrate the manner in which the standard was interpreted and explained to the participants by Jenny. She commented: “Note the achievement and merit criteria are the same for a particular aspect like planning”, and drew attention to the sole difference between merit and excellence grades for the same aspect, that is, the use of the descriptor “sufficient” in the excellence section. When Jenny queries Sam about the meaning of ‘sufficient’ in this context he was able to explain: “More details, so it’s got instead of just the average it’s got more information”. The frequent questioning and probing of students’ understanding about the requirements for achievement by Jenny, and the nature of the student responses as illustrated by Sam, revealed that in these early stages students were being made more aware of the procedural knowledge required to meet the achievement criteria of the standard – they were being coached in how to ‘deliver their lines’. In another instance Jenny asked Mitchell what the phrase “provide a comprehensive evaluation or discussion of the investigation” means for achievement at excellence level. After a short series of interactions with Jenny, he responded with the explanation that it means giving “quite a detailed summary” of the investigation.

Following the shared reading Jenny directed students to an exercise in the student workbook where in pairs they were required to work through a page of terms related to investigations, discuss the meaning of each term and fill in definitions for each. She told them to check their definitions with those in a glossary at the back of their workbooks. A short period of discussion followed and then most students in the class worked silently on filling in the definitions. Peter and Martyn were focused on this task, but Eddie and Mitchell spent little time on it and were soon disengaged. Mitchell had forgotten his workbook and was directed by Jenny to share with a partner, but the sharing did not eventuate. Jenny intervened at this stage to give students further guidance, pointing out pages further on in the workbook where a planning and reporting template modelled on that used for the formal summative assessment is depicted: "I want to show you where the terms fit into the template". This time Jenny was ensuring students were learning the 'right lines' and emphasising the means by which these 'lines' could be delivered to meet the achievement criteria of the standard: "Remember and learn what's going on in the boxes. It's all on the Internet … no excuse for not knowing what to do". After asking several students for their definitions Jenny writes up on the whiteboard the 'correct' definitions from the teacher's guide, which most class members record in their workbook or notebooks. As she provides these definitions she frequently refers back to the achievement criteria and planning templates requirements, highlighting when a particular term will be needed and sometimes seeking students' feedback on understanding: "How many variables do you need for excellence?" Most of this teaching episode involved the teacher either eliciting information from students through questioning (to monitor their comprehension of the requirements of the planning and reporting template) or giving information; and students listening, answering questions and taking notes.

In the second lesson work on the standard did not get underway till the latter part of the period when Jenny delivered a brief teaching episode on the process of dilution. She explained that this was a skill the students would need for their formative assessment, which was happening in upcoming lessons. Jenny taught the concept of percentage concentration and demonstrated how various concentrations could be obtained using aqueous potassium permanganate solutions. The students then had five minutes in their groups to practise diluting their solutions to specified concentrations. My student group participated enthusiastically and, although there was some off task behaviour and horseplay, they were able to achieve some success in this task. As homework for the third lesson the class was asked to do a planning exercise from the workbook (Cooper, Hume & Abbott, 2000). This homework task initially involved students planning an experiment to determine "which Bunsen flame, blue or yellow, heats water the fastest?" In the workbook they had to fill in a planning template, modelled on those used nationally for the formal summative assessments.

At the beginning of the next lesson to complete the homework planning exercise, the students were required to work in pairs to peer assess each other's work, using the assessment schedule provided in the workbook. Jenny gave them some pointers about using the marking schedule before they started: "You need to be specific when marking", so when describing details about how a variable such as the amount of water is to be controlled, she told them it is not sufficient to use the term "same", they needed "same height". When the peer assessing began, my participants were slow to start and Eddie, who had not done the homework, was told by the teacher to do it in class. Once underway, they started the task in silence, but it was not long before conversation began as they sought clarification and reassurance from one another about aspects of the marking. For example, Mitchell revealed his interpretation of repetition: "You've got to do it 60 million times". Several minutes later Jenny intervened, commenting to the class that if "they miss part of the achieved, they miss achieved". Martyn realised that even if he met the merit and excellence criteria he would fail the whole standard if he didn't meet all of the achieved criteria. He expressed his frustration labelling it "stupid". From this point on the group became disinterested and very soon were off task. Shortly afterwards when Jenny sought feedback from the class on achievement rates, 10 students had achieved the standard at achieved level while none had achieved with merit or excellence.

When Jenny called for more work to be done by the class to improve overall achievement, in response to the homework results, the class expressed a general consensus that the assessment schedule for the homework task was hard to follow. They thought the task was difficult to mark because there was not enough detail in the schedule. A short whole class discussion followed, prompted by student queries about how the teachers in the school make assessment decisions for the standard. Jenny informed the class that “science teachers do it in consultation with a teacher in charge of checking consistency”, and reassured a student that the process was the same for all students in other classes.

4.2.4.3 The lead-up to the dress rehearsal

Following the homework session in the third lesson Jenny introduced the instructions for the practical investigation that was to the subject of what was termed the ‘formative assessment’. The purpose of the formative assessment was to give students the opportunity to experience and learn what they had to do to meet the criteria of the standard at achievement, merit and excellence levels – it was a form of mock summative assessment, similar in most procedural ways to the final assessment event, except for the actual scientific phenomena being investigated. The instructions for this ‘dummy run’ investigation, into the effect of acid concentration on the rate of reaction between magnesium metal and hydrochloric acid, included experimental details and relevant background science concepts (reactions of metals with acids, and rates of reaction) – see Appendix J. The students already had practical experience of the background science because this formative practical followed directly on from the work done in class for Science Achievement Standard 1.4 *Describe properties and reactions of groups of related substances* (an externally assessed standard). This standard required students to know the products of acid-metal reactions, including those of the reaction of magnesium metal with dilute hydrochloric acid. Thus the students had had previous experience with the specific reaction under investigation in the formative assessment, including factors that affected its rate of reaction.

The background science notes supplied to students for the formative assessment contained a summary of key scientific concepts related to the investigation (see Appendix I). These concepts included: a description and labelled diagram of the magnesium/hydrochloric acid reaction; a definition of reaction rate and methods for measuring rate using a similar reaction to the magnesium/hydrochloric acid reaction; an account of the collision theory of matter; and, methods for increasing rate of reaction explained in terms of the collision theory. The students had also briefly experienced the theory and practice of dilution in lesson two, just prior to the formative assessment.

During the remainder of this third lesson Jenny lead the students through the instructions for the practice investigation, which was to begin in the following class (lesson four) and extend over four lessons. She focused on the planning stage, which was to be done individually during lesson four. Jenny modelled the planning process by systematically addressing each planning phase, and interacting with class members in the whole class forum to elicit the details required for each step. Using probe questions like: “How will you see that a reaction has happened?”, Jenny helped students to select and contribute relevant concepts from their existing knowledge frameworks, such as “bubbles will be coming off” to utilise in their planning. She often prompted students to make links with prior experiences by encouraging them to recall similar situations, using phrases like: “Remember when we did?” and “How can you measure gas?”. In other instances she encouraged students to evaluate contributions by making comments which pointed to strengths and weaknesses in aspects of planning – when a student suggested using a Bunsen burner to control the temperature at which a reaction occurred she asked: “Do you think it would be easy to use a Bunsen to change the temperature of the water?” Sometimes she took time to quickly revisit and revise concepts and skills, like using diagrams to show the equipment and techniques involved in collecting a gas, or demonstrating the use of a mortar and pestle. When concluding the lesson, she reminded students of the key components of fair testing, and stressed the need to be well organised to achieve excellence. For homework students were asked to do exercises from the workbook that involved processing data (tables, averaging and graphs).

4.2.4.4 The dress rehearsal

From my observations of these first classroom lessons, and my first interview with the study group, I was unable to detect any clear indicators of the nature and extent of their understanding about scientific inquiry. However, during the following four lessons of the formative or mock assessment (lessons 4 through to 7), when students actually planned and critiqued one another’s plans, and carried out investigations, indications about learning began to emerge. These lessons proved to be the first opportunity for me to gauge students' degree of mastery of scientific concepts, skills and procedural knowledge from their activities in class, and any learning progress that they may be making. For ease of purpose I've chosen to set the scene first by describing the sequence of classroom events that occurred during these four lessons first, then returning to give a more holistic account of the responses and reactions of my study group as they experienced these events. The sequence of events within each of the four lessons follows.

After a brief introduction in Lesson 4, where they are given the instructions for the investigation including the template for the report, the students began planning the experimental part of the investigation individually in class. An identical report template was used across all the Year 11 science classes in the school, and was modelled closely on that provided in NZQA guidelines. At the close of the lesson the teacher collected in the plans. She read the plans after the lesson but did not mark them or give any written comments.

Lesson 5 began with some global feedback and feedforward from the teacher on the class's planning to date. For example, she commented that; "Scientists alter their methods so it's okay for you to do it in your work", and went on to briefly cover extra teaching points and give tips where she felt planning aspects needed extra thought. The students then moved into their groups to make collaborative planning decisions and carry out the experimental work. The teacher initially left the students to work on their own (as they would be doing in the summative assessment), but later in the session she interacted with some groups when equipment and/or chemical supply issues arose. Towards the end of the session, she called for whole class attention to pointers about the recording and processing of their data, using the workbook (Cooper et al., 2001) as a reference, and to happenings in the next phase of the formative exercise. Students were reminded that the report write-up in the following session was individual, and that they would need a record of the data. Jenny suggested that in the final minutes of the lesson they "use the brains of the group while they are together so you've got ideas for tomorrow". Again Jenny collected up their planning and recording sheets at the end of the lesson.

At the beginning of Lesson 6 Jenny returned students' planning sheets and spent approximately five minutes revising key requirements for the final phases of the write-up. She drew students' attention to the collating, analysing, and interpreting of data and to the evaluative section with its links to the background science. To emphasise the data processing that needed to occur Jenny referred students to the homework exercise set several days prior from their workbook, and reminded them of important graphing items such as even scales on axes, titles and use of rulers. The students were able to refer to their notes and for the rest of the period they individually wrote up their reports. Jenny made herself available to answer students' questions mostly on an individual basis, but on one occasion she stopped the class to stress the need to identify trends in data when writing up the interpretation section and to revisit the concept of systematic errors. Once completed the students handed in their reports to Jenny.

The main activity in Lesson 7 was the peer assessment of the students' reports. Jenny had assessed the reports overnight but no grades or written comments were evident when she handed them back to the class. She distributed the assessment schedule (see

Appendix K) to the class and spent approximately 10 minutes walking the students through it, giving them very detailed feedback and feedforward in relation to aspects of the schedule. In this time she modelled how to assess the report, providing instances of evidence that demonstrated achievement of various criteria. Jenny then instructed the students "to act like a professional teacher and mark the work". The majority of the class began the assessment activity but a small group who had missed earlier lessons worked with Jenny receiving catch-up information. The students were to do the assessment on their own but very quickly began consulting one another as they progressed through the task. Later Jenny directed them to work in pairs to discuss and justify their marking, and during this period students were able to approach her at the front desk with queries about the marking. Only a few students approached her with questions about the assessment.

4.2.4.5 Lessons from the dress rehearsal

As the formative assessment exercise progressed I was able to observe the group members using and applying many fair test terms and concepts appropriately, in their discussions amongst themselves and with their teacher Jenny, and in their actions. For example, they could readily identify relevant variables, suggest suitable ways to control variables, and distinguish between independent and dependent variables. In their planning they appreciated the need for a suitable range of values for the independent variable, for repetitive measurements and for taking into consideration sources of error. They could also recall and apply basic and appropriate data gathering and processing skills, including observing, measuring time using a stopwatch, recording data in tables, averaging and graphing. As they planned the formative investigation, this episode from one of the lessons will show that my study group were also successfully managing to integrate some of their knowledge and understanding of pertinent background science concepts and skills with their knowledge of how to do fair testing. It will also show that the group, and individual members, encountered various problems when executing their plan for their investigation during the practical part of the formative assessment.

As part of the experimental design process Martyn, Peter, Mitchell and Eddie had to make a decision about quantities of reactants. They had identified acid concentration as the independent variable and decided to test six different acid concentrations of the acid. They also realised that equal amounts of magnesium ribbon were needed in each experimental run to ensure a fair test, so a quick decision was made to divide their original sample of magnesium ribbon into six equal pieces. Without any trials, several group members went ahead and started the experimental runs only to discover that no one in the group was timing the reaction. A second run using the same concentration of acid was started, but again they realised no one was timing. Martyn made the decision to take responsibility for timing and a time was successfully obtained on the third attempt. After several more runs, using different concentrations of the acid, Peter suddenly realised no one had recorded the results so he did it retrospectively on paper. Their final run took so long for the reaction to go to completion that all the group members became disinterested and went off task. When they did finally re-engage with the task, Martyn commented the results didn't make sense and decided the experiment using the first acid concentration needed to be repeated. At this stage the group became conscious that they had run out of magnesium and would not be able to obtain sufficient results for their graph. Eddie alerted Jenny to their plight, and she gave them extra magnesium. The students were eventually able to obtain results for five different acid concentrations, not six as originally planned. Their sixth acid concentration was so dilute the reaction did not go to completion in the time available to them. As a result of their failure to foresee and address potential design problems through trialing, the students consequently experienced great difficulty performing the practical side of the investigation and obtaining sufficient data.

On reflection, it became clear that most of these problems arose because the students were doing insufficient forward planning. They tended to neglect the finer points of experimental logistics like the need for: teamwork and delegation of responsibilities; realistic levels of repetition; accuracy in measurement; construction of tables for data recording prior to data collection; appropriate units in tables and graphs; and, the use of tools like a ruler for drawing up tables of data. My study group members appeared not to appreciate the extent and degree to which they had to plan, and consequently they often did not become aware of the shortcomings in their design until they were well into the data gathering stage.

The group's actions in the formative assessment also pointed to apparent gaps in their procedural knowledge regarding the rationale for conducting trials, and the role trialing plays in determining the finer details of experimental design. I witnessed their 'headlong rush' into the formative practical, after minimal consultation as a group on the best plan of action for the experimental work, which led to breakdowns in teamwork at critical times. The students' haste to get underway, and their failure to fully consider all factors affecting the experiment, appeared to indicate their lack of appreciation of the preparation that goes into planning and the reasons for trial runs. It may also have reflected their anxiety about the assessment, particularly the time restraints. The instance quoted during the formative assessment session illustrates through the students' behaviours how this breakdown occurred and the impact it had on the outcomes of their investigation.

The students' written assessment records for the formative investigation reflected the knowledge of fair test procedures that they had displayed in class. To achieve the standard students had to gain a minimum of 'achievement' in all three sections of the report – the planning section, the results and the conclusion. If they failed to reach 'achievement' standard in one of these sections then they did not achieve credit for the whole standard. It is important to note that this formative assessment was graded by student in a peer assessment exercise, and there were a number of instances in my participants' work where the peer assessors' judgements were not aligned with the evidence required in the assessment schedule. Where these non-alignments occurred in the groups' reports I have attempted to describe the discrepancy and indicate how the judgement on learning might change if the evidence was more closely matched to the assessment schedule. The peer assessors also did not have the discretion as their teacher did to award 'on balance' assessments.

The group's lack of attention to detail and rigour in planning and performing the fair test was also revealed in their written responses, but not to the same degree that these gaps in procedural knowledge manifested themselves in the practical sessions. Of the four students making up the original group, Martyn achieved the highest grade for his formative assessment, managing to reach merit level. While he was able to achieve at excellence level in the results and conclusion sections, his planning section was judged at merit level, that is, feasible but lacking in sufficient detail to be independently followed without further clarification. For example, in his original plan he made no mention of the amount of magnesium ribbon to use in each reaction, or how to dilute the acid to the required concentrations. He did correct these omissions later, after feedback from his peer assessor and Jenny when she fed back to the whole class on common strengths and weaknesses. His plan also required a total of 18 timed reactions, which was ambitious considering the plan did not suggest running a number of reactions simultaneously. When I examined his results table more closely, I discovered he had used two different units (minutes and seconds) within the time column, but on the accompanying graph he had corrected this using only one unit (minutes) on the time axis. In the evaluation section Martyn correctly interpreted the findings, drew an appropriate conclusion to the investigation from the findings and attempted both parts of the evaluation section – an analysis of the success of the investigation including any modifications if repeated and why; and the linking of the findings to science ideas. He was able to list most of the difficulties experienced by the group carrying out the plan but only suggested one means of overcoming these difficulties (repeating the experiments for greater accuracy in the results). He was unable to suggest any systematic errors or make the links between the findings to the relevant science ideas (collision theory).

Peter, Mitchell and Eddie failed in their written reports to achieve the standard. Like Martyn, their planning sections lacked the detail to make their plans workable. For example, Peter when describing the amount of magnesium ribbon required in each reaction wrote: “Get the right amount of magnesium ribbon for each trial from the amount given at the start (6cm)". He also made no mention of the requirement to progressively dilute the supplied HCl to achieve the desired concentrations. Mitchell described a dilution process in the first steps of his method, but omitted to mention the acid. His instruction to "repeat each reaction five times" gave a total of 30 reactions to time – from the method as written the reader would take it that these reactions were to be carried out in sequence and not simultaneously. Eddie was given 'non-achievement' for this section because his aim was incorrect, but the rest of his plan had sufficient detail to be feasible. All three 'achieved' the results section with Mitchell gaining excellence and Eddie merit. However, when I looked more closely at their papers I noticed they had used two different units when recording results, and Eddie did not use a table – according to the assessment schedule a table was required for minimum achievement, and he should therefore have been given 'non-achievement' for his results section.

In the conclusion section of his report Eddie made a series of statements that alluded to the findings, but all lacked the precision required by the assessment schedule. His peer assessor granted him an achievement grade, despite his failure to provide a correct description of the trend in the data in his interpretation section: "There was an enormous difference between 100% and 0%. The graph curved nicely". However, he did suggest an appreciation of the trend in his conclusion statement: "We found that the 100% acid burnt the magnesium strip away faster than the 0%". It seems likely his peer assessor gave him 'achieved' on the basis of this information. While Mitchell's comments showed greater clarity, with conclusions based on the findings and links between his findings and the background science of collision theory, he was not accurate in his interpretation of the results. He described the trend as "the higher the concentration of HCl then the faster the magnesium reacts" when the marking schedule called for a trend between the acid concentration (the independent variable) and the time (the dependent variable) to be described. Since this was the procedural evidence required for achievement he received 'non-achieved' in this section and failed the whole standard. Had Jenny been marking the paper it is likely, given the several merit and excellence points Mitchell was able to provide, that he would have been given at least 'achieved' for this section. Mitchell did not comment on how well his group had performed the investigation or suggest changes.

Similarly Peter successfully met most of the criteria for merit and excellence in the final section. He identified several of the significant difficulties experienced by the group carrying out the practical, including the failure to delegate tasks and the need for more results to form a conclusion, and he successfully linked experimental results with relevant science ideas. His account of systematic errors suggested that he did not have a strong grasp of this concept. He credited the disparity he perceived between his graphed results and those he expected to systematic errors, indicating his understanding of this concept was more closely linked to the scientific concept of 'outliers' (pieces of data that do not fit the general trend of data) than it was to the effect of regular procedural errors on data, such as the small time delays when humans activate stopwatches. Like Mitchell, one inaccurate statement cost Peter achievement in this section - it was the inclusion of the phrase "the more concentration the slower the reaction time" in his interpretation of the results that prevented Peter meeting the essential achievement criteria. Consequently his peer assessor judged that he had not achieved the criteria required to gain credits for the whole standard.

Although I did not observe Sam performing his formative assessment I was able to examine his written response. On this basis Sam performed the formative task at achievement level, and there were many parallels in his work with that of Martyn, Peter, Mitchell and Eddie. His planning section lacked the detail and clarity to make it workable, but he did mention control of an extra variable (the reaction temperature using a water bath) not covered by any of my group. The method for diluting the acid was not included, and it was unclear in Sam's method if there was to be repetition of experiments - the results section showed only one run for each different acid concentration (six timed reactions in total). His peer assessor failed to pick up his incorrect identification of the size of the magnesium strip as the dependent variable. Data was placed in a partial table, but again like his peers he used two units (minutes and seconds) within the table. In his conclusion section some of Sam's comments hint at appropriate scientific understanding, but in these instances he is unable to express himself with sufficient accuracy and coherence to make his meaning clear. For example, in his interpretation of the results he writes: "The lower the reactant the longer it takes the magnesium strip to dissolve", where it could be assumed from his later comments in the conclusion section that he meant to say "the lower the concentration of hydrochloric acid the longer it takes for the magnesium to react". He is able to list areas of the investigation that produced difficulties, including the discovery that trying to maintain a constant temperature for the reaction was too time-consuming so his group abandoned this component of their experimental design. Sam did not suggest a solution for this problem, but he does describe how the difficulty of magnesium sticking to the sides of the test-tube can be solved by always using dry test tubes. He makes a credible link between his findings and the collision theory, but does not identify systematic errors.

It was at this point, about two thirds of the way through the teaching and learning sequence, that I held my second interviews with the teacher and with the students.

4.2.4.6 The director's progress report

Jenny's overall estimation of the class performance in the mock assessment was favourable: "Not too bad". Basing her judgement primarily on information she had obtained from marking their formative assessment scripts she commented positively about certain aspects: "Most of them have written really nice methods which were definitely near it [i.e., excellence level]. They were probably better methods than I have seen, as far as students having their first go at this". This assessment information also pointed out some gaps in their learning: "Their variables, they didn't do that very well".

While happy with the performance to date she was particularly taken aback by one incident in class, that is, the students apparent 'refusal' to keep quiet during the peer assessment. Despite several requests to do the task quietly and independently, she had been surprised by the students' persistence in talking to one another. In hindsight, she now realised "it was to do more with the fact that it was quite difficult for them and they were asking what was happening all the time, and they actually needed to talk to one another". She recounted her own experiences: "I thought back to my NCEA days, when we first started marking the scripts as teachers, and we were used to marking. We actually talked to each other a lot as we were marking them. What about this point? What about this?" Jenny recognised the value of peer assessment as a learning strategy: "For next time I would definitely just let them go ahead and chat".

4.2.4.7 The actors' opinions on their performance

In their interview after the formative assessment, the group members in attendance (Eddie was absent) found it difficult to identify specifically what they had learned in the investigation unit. They chose to use general descriptions like "how to do science experiments well, draw up results and all that discussion", and "the order to write them in". Martyn commented that "I quite often know what they are already talking about already", intimating that he was already familiar with much of the material covered in class, and therefore felt he was not learning many new concepts or skills. They did feel doing practice investigations helped their learning. Other strategies that helped included listening, trying to stay on task, more time and more working in groups "because then you talk about it". Martyn and Peter thought "catchy little songs" and "jingles", like that used to remember elements in the periodic table: "Hi, he likes beer but in cups not over frothing" might be good to remember the science needed for the evaluation section because "it is stuck in your head for the whole day and then you know". Peter intended spending time revising before the summative assessment studying "the brainstorms, the notes and the exercises in that book [the student workbook]".

The students appeared to be more certain that their views about NCEA were changing as a result of their experiences in class. Martyn's concerns about NCEA had lessened: "Its not so bad … NCEA is good the way you get several tries, so if you fail once try again." He was not as concerned about getting merit and excellence grades, he just wanted to pass: "If I get achieved I am happy with that" - the others agreed. Peter's worry about NCEA had intensified, however, because the formative assessment had not gone well for him - he had not achieved the standard. In his final section he had made an error using the term ‘slower’, when on the basis of his results he should have used the term ‘faster’: “I knew all the first okay”, but in the last section he said he “didn’t put in some word I was meant to”. After this experience he felt he was not going to do well in the summative assessment, unless “I get taught heaps more about it”. He was not confident that much more teaching would happen. Mitchell, who had appeared unworried about his learning for NCEA in the first interview, was now feeling that in NCEA “I am not going to do terribly good … it’s so much pressure”. He did express some relief at achieving at achievement level for the formative assessment.

4.2.4.8 Last minute stage preparations

In an earlier departmental meeting, Year 11 science teachers had been briefed about the summative assessment and given explicit instructions (oral and written) about how the assessment was to be conducted. For moderation purposes the teachers were to closely adhere to these instructions to ensure that no students were disadvantaged. In the meeting, the teachers discussed with the teacher-in-charge of the Year 11 science assessment, what information could and couldn’t be taught, said and given out. Thus all students were to have access to the same background science concepts and skills through pre-teaching of this material by classroom teachers. The conditions under which the investigation was to be conducted in class were also to be identical, including the time allocated for the various stages of the assessment, and the retention of students’ written work and notes between lessons by classroom teachers. The teachers received the common assessment schedule, designed to provide consistent marking across all classes. Jenny followed the guidelines closely as she prepared students for the investigation beforehand, supervised the practical sessions and assessed their written reports.

Between lessons seven (the end of the formative assessment) and lesson eight (the lead up to the summative assessment) Jenny had required students to do a homework exercise from the student workbook (Cooper, Hume & Abbott, 2001). In a scenario situation students were to participate in a simulation of a practical investigation. The scenario involved two people investigating which fabric type would make the warmest jacket. Other information supplied included:

- an hypothesis and some tips, in the form of questions, about things to consider before writing the plan
- a description of how the two people carried out the practical side of the investigation
- a description of the data they collected, for example: "The five melting/burning times for polar fleece were 1.7 s, 1.5 s, 1.3 s, 1.6 s and 1. 4s."
- a graphing grid, and
- tips, in the form of questions, for interpreting data.

In the exercise students were first to write a plan using a template similar to that used by them in their formative assessment. They were then to tabulate, graph and interpret the data provided. Finally they were to discuss and evaluate the experiment in terms of how the experiment could be improved. An assessment schedule for the investigation, based on the NCEA model for AS1.1 was provided on two pages immediately following the scenario in the workbook.

At the beginning of the eighth lesson Jenny informed the class that the summative assessment would be an investigation into pendulums, specifically the relationship between the length of a pendulum and its period of swing. She instructed the class that, as part of the preparation for the summative assessment, she would give them "formative practice" with actual pendulums later in the period, but in the first segment of the lesson they were to pair up with their neighbour and peer assess their homework task (the scenario investigation into which fabric type would make the warmest jacket) using the assessment schedule provided in the workbook. Before students began this peer assessment task Jenny talked them through the schedule in the workbook, drawing their attention to elements within that schedule, which were also applicable to the summative assessment investigation. Typical of Jenny's hints were: "Give lots of detail for excellence", and "the practical tomorrow requires you to change the length of the string of the pendulum and as you change the length of it you'll get a different result. Now you'll need to try quite a few different lengths. How many lengths will you use?" A student replies "six" and she agrees, adding that six repeats of each length will also be needed: "Remember the rate of reaction experiment".

As the peer assessment got underway Jenny commented; "You can talk together when you mark because I saw last time that you found it hard to stay quiet, you needed to talk. Try marking it quietly first, then discuss". None of my study group were sitting together, so I chose to observe Martyn and his neighbour assessing one another's work from the schedule. They immediately began discussing the marking, focusing on individual points such as feedback: "Your discussion sounds like a high power scientist"; omissions: "You forgot types of fabric"; further clarification: "How would that improve it?"; and, justification of their responses: "I said this is how you insulate. Remember I said you had to wrap it up?" Martyn and his neighbour initially gave their full attention to the task, but after a time frustration set in. At one point Martyn exclaimed "I don't get this!" and his partner replied "Neither do I". At this point they began to disengage from the task. Scanning the whole class it became clear that some students had not completed the homework task, and they were therefore not participating fully in the peer assessment task.

After the peer assessment exercise Jenny turned her teaching focus to the summative assessment. To allow students to prepare for the summative assessment, which was scheduled for the next day, she first handed back their formative assessment reports so they could revise overnight. Then she taught key science concepts associated with the operation of a pendulum, referring the students to notes for the summative assessment which she had also handed out – these notes included a brief account of the background science with a labelled diagram of pendulum components and the task instructions (see Appendix L). Jenny set up a model pendulum, which was attached to the support strut of an upturned stool, and demonstrated various pendulum concepts and skills from the front of the room on a model pendulum. She then reiterated many planning aspects in the context of the pendulum, focusing student attention through questioning and discussion on the right terminology to be used, where to locate equipment, what is to be measured and how it should be processed, the importance of pre-trialling and working as a team. The students were then given 10 minutes in their groups to do trialling of possible methods.

Although my study group had not felt in their second and even third interviews that much learning had occurred in the formative assessment, I did observe them paying more initial attention to trialling, and to the finer details of some aspects of planning for their summative assessment. Mitchell was absent for the trialling session, but the remaining group members used their time more productively than they had done in the formative assessment. They took the opportunity in the trial time to debate various design points, and to address issues to do with technical details such as ensuring their pendulum swung freely. From their comments and actions in addressing the experimental design in the trials it was clear that group members had a stronger appreciation of the procedural knowledge detail required to keep the testing fair. For example, Peter kept the group focused on the independent variable (the length of the pendulum) and the systematic manner in which they intended to vary it, while Martyn and Eddie took care to establish accurate methods for counting and timing pendulum swings to minimise systematic errors.

Team-work was improved in the trialling, but unfortunately the group experienced technical problems with the swing of their pendulum, both in the trialling period and as it eventuated in the summative session. First, due to knot tying problems, they experienced difficulty changing the length of the pendulum in a systematic way (10, 20, 30 cm and so on) as they had planned to do. Then they had trouble obtaining the planned number of consecutive pendulum swings due to the pendulum colliding with the leg of the stool. These difficulties caused them immense frustration and were not solved in the trailing phase. When Jenny calls a halt to the trialing period Martyn exclaims: “Oh man!” Despite many attempts they were unable to solve the problem satisfactorily, and these same aspects of the experimental design returned to haunt them in the actual data-collecting phase of the summative assessment.

After the trialling episode in lesson eight Jenny resumed providing relevant information and direction about how to approach the summative assessment. Through her focused questioning such as “which one is the independent variable?” and “what factors/things might affect your experiment, for example, if you did it outside?”, Jenny enabled students to identify pertinent information about the investigation, including the identity of the independent and dependent variables, the type of graph needed, how to produce a line of best fit, appropriate labelling of graph components such as axes and sources of possible error. She also anticipated certain requirements and alerted students to possible solutions should the need arise: “You may have to use two results columns in your table because you are graphing for one period”. Students’ questions were also answered; one student asked: “Does it matter if you swing it (the pendulum) horizontally, or the height you release it from?” Jenny responded saying that “the height of release does not affect the length of the period, only the length of the string”, but she was unable to explain why. She also points out that “it doesn’t matter if you can’t get even lengths for your pendulum string, so long as you have even spaces for units on length axis it will work out”. Jenny stopped short of telling them whether the line graph was a straight line or curve “because I think I’m teaching you to do it anyway”. At the close of the lesson she collected in their instruction sheets for safekeeping and reminded the class to use their homework exercises for revision purposes.

4.2.4.9 Coping with final night stage jitters

Lesson 9 was Act One of the summative exercise – the planning phase - and students carried this out individually. Jenny gave last minute reminders about procedural points to do with the planning, then assumed an examiner's role telling the students that the assessment was "closed book, and not a learning exercise. I cannot answer questions unless it is really important". Sam immediately volunteered that he and some other students "won't be here tomorrow, so won't actually see what happens". Jenny reassured him that she would spend time with them when they returned to class for the report writing to explain what happened, and that they could also talk with their group for five minutes. While other members of the class began their planning work, Mitchell was called up by Jenny so he could be briefed on the task, since he had been absent the lesson before. Once briefed, Mitchell joined the rest of the class to write up his plan. As the class worked Jenny responded to a number of individual students with questions. In quiet one-on-one interactions Jenny often deflected the question back onto the student by refocusing their thinking: "Look back to the original sentence at the beginning of section two". Martyn, Peter, Mitchell and Sam appeared engaged in the task for most of the period but Eddie was often not engaged and often appeared unsure how to do the exercise. At the end of the period Jenny collected in the individual plans. They were returned to students in the next lesson, but no marks or grades are evident on the plans

The next day Act Two began - the data-collecting phase. My group initially appeared to be more purposeful in their approach to the investigation than during the formative investigation, with Peter and Martyn collaborating closely in the setting-up of the apparatus and trialling of the pendulum swing, while Mitchell and Eddie drew up their tables. They had made modifications to their apparatus utilising a retort stand and clamp instead of an upturned stool as support for their pendulum. However, when I asked Martyn what plan they were using he was unsure – they had not decided at that stage on a specific plan. The discussion that followed between Martyn and Peter, about the number of repeats to be performed, confirmed that some planning decisions were still to be made. When Mitchell and Eddie joined in the discussion the group could not agree, and Jenny's advice was sought. Jenny did not provide an answer, but fielded the question back to the group. Finally, after consulting with others out of the group, my group came to a consensus that they needed to do six repeats for each pendulum length.

Shortly afterwards Martyn and Mitchell, who were then working on the apparatus together, struck a problem. They realised that the pendulum was not swinging from a fixed point, so they experimented with different forms of attachment. As they tried various methods Eddie began to express frustration saying "start again", and "hurry up, hurry up!" As in the formative assessment it became evident again that the group had failed to assign individuals to tasks: "Whose got a watch?" asked Eddie, "I've got a watch … need a ruler!" replied Martyn. "The teacher took it off us" said Eddie, so Mitchell goes off to retrieve the ruler. Despite some group dissatisfaction with the set-up, the experiment began. Almost immediately they realise no one was timing, so they restart with Peter operating the stopwatch. He successfully timed the first run, and Peter quickly sketched up a table on a piece of loose paper to record the results. The next few runs are recorded, but Eddie was worried about how long it was taking crying "hurry up, hurry up!", and he pointed to the set-up of a neighbouring group suggesting "do it like they did". The group began to swap around roles with Eddie, then Peter, swinging the pendulum and Mitchell recording and calculating average times.

Shortly afterwards, the technical difficulties they had experienced in the trialling resurfaced. The pendulum began to swivel in its action and after 6-7 swings the pendulum bob would collide with the support leg of the upturned stool. When they failed to record successive runs after several attempts panic began to set in: "Hurry up! We won't get it done!" said Eddie. Further time was lost as they struggled to adjust the length of the pendulum to an exact measurement. It appeared they had not heard or comprehended Jenny's comment the previous day that the length of the pendulum did not have to vary in a systematic way. Martyn took over the task of adjusting the length exclaiming: "This is out of control! It's supposed to be 20! This is stupid!" When Peter finally suggested that "it doesn't have to be exactly 20 centimetres", Mathew replied "yes it must, it's supposed to be a fair test". Eddie became exasperated; "Martyn, stop fiddling with it. We're never going to finish it if you keep fiddling with it!" Martyn finally adjusted the pendulum to 20 centimetres and the experimental runs recommenced.

Unfortunately maintaining the pendulum for 10 consecutive swings remained a persistent problem, and the group's levels of anxiety continued to grow, particularly when they witnessed other groups in the class successfully completing their data collection. Amid the growing frustration and confusion Peter took on a calming, reassuring role, telling his fellow group members not to give up hope: "We can do it!" Martyn commented hopefully "do you think we could argue that we started late?"

Eddie sensed from their previous experience that their method was not working and that they should "stop it now and start again". He wanted to alter the method but the others were determined to carry on with the existing plan. As pressure mounted to complete the experimental work within the time allowed even Peter began to worry: "There's no way we're going to get this done!" He took on a leadership role urging Eddie to help Martyn adjust the pendulum length again.

Within the closing stage of the practical session the group scrambled to complete and record sufficient runs for their data processing and interpreting phase. The four group members frequently interchanged roles as they each took it in turn to record their own copy of the results (which they needed for the write-up in the following session). All other groups had finished their data collection and were listening as Jenny covered points for the write-up. Martyn, Peter, Mitchell and Eddie continued operating their pendulum and consequently missed hearing what Jenny was saying during her briefing. In their rush to finish confusion set in: "Is this the third or fourth one?" asked Mathew who was recording and calculating. When the pendulum continued to collide with the support arm Peter commented "You'll have to estimate", while Eddie was convinced they should "make up the rest". Mitchell agreed: "Lets make up the rest, and take 16 seconds as the average" and Martyn confirmed "it will still give us our results". Each group member had a complete set of written data by the end of the practical. Jenny allowed the class to view the background science notes before the end of the period before collecting in all papers to retain overnight.

The next day the students were to perform the Final Act by completing the write-up individually under exam conditions - they were not to consult their background science handouts. While the class was settling Jenny briefed two students who had been absent the previous day at the front of the room. I noted that later in the session she withdrew a further two students out of class for several minutes. They were asked to brief the absentee students on the happenings of their group investigation the day before, and to provide them with data.

Once the class was settled and before the class started the write-up, Jenny spent the first 5 – 10 minutes of this summative session guiding students through the requirements of the last two pages of the assessment template. The students were reminded to finish their data table first, to remember graphing skills: “Check back to the purpose so you have the right heading. Look for the trend on the graph. Is there a pattern?”, and to “relate their conclusion to their experiment”. She focused on the evaluation section, encouraging students to recall aspects they had missed in the formative assessment, like reliability of data, systematic errors, improvements to the method and linking of the findings to the background science. She answered a student question about systematic errors and provided graph paper to those students requiring it. Several students, including Peter, requested rulers and Jenny obliged. The class was given 45 minutes for the write-up, and all my study group, except Eddie, remained on task for that time. Eddie had significant periods of time where he appeared unable to make progress with his report, but he would return to writing before experiencing another lengthy period of apparent inactivity. He also handed in his paper about 10 minutes earlier than most of the class. Jenny allowed Martyn to work for a few minutes after the end of the period completing his report.

4.2.4.10 The director as critic

Jenny critiqued the students' performances, marking their summative scripts with the same assessment schedule as other teachers in the science department. For moderation purposes, the Year 11 teacher-in-charge check-marked a sample of scripts from across classes. Jenny gave the scripts back in class, but before students received their scripts she gave five students in the class a reassessment opportunity. These students were very close to achieving at an achievement level or merit, but were just missing a point. Jenny called students up in turn to discuss their script and revisit the pertinent section containing the point in question. If students could self check their 'error' in this discussion, they were given the higher level of achievement.

Once the reassessments were over Jenny asked the class to carefully check the marking of their scripts, as she gave feedback on their performance in a detailed, step-by-step description of how the task was assessed. This process was similar to the method she used to provide feedback after the formative assessment task was marked. For example, she identified features from the assessment schedule that must be present for achievement, and highlighted common errors e.g. graphing. On some occasions Jenny took the opportunity to illustrate how these common errors could be corrected. Going over the results section she commented: "Graphs let a lot of you down". She then took the opportunity to model how to get a line of best fit. Over all, the detail Jenny provided in this feedback session was more specific than that given in the formative assessment, and much seemed aimed at justifying the marking decisions. She also explained the principle behind the holistic approach the teachers had adopted in their marking, where small errors were disregarded if the overall standard of the work was high. Teachers in the department had been given the discretion to use their judgement to award 'on balance' grades for situations like this.

All students in my study group showed improvement in their achievement, although these improvements were not necessarily reflected in their global grade for the standard. As illustrated below, when I looked more closely at their written reports there were instances, which demonstrated significant learning gains by the students in this case within certain report sections. However, these improvements were not sufficient to warrant a whole achievement level shift upwards, according to the assessment schedule.

Martyn's summative results revealed noticeable improvements in his planning and method section. His description of the method was more detailed and informative than his formative attempt. Improvements included a more logical sequence of steps, a labelled diagram, and fuller accounts of appropriate means of controlling variables. Despite these clear improvements his plan was not considered detailed enough to be considered workable, that is, a method that could be independently followed without further clarification to obtain reliable data (see marking schedule in Appendix M). In his results section Martyn again achieved at excellence level with displays of the collected data in appropriate tables and graphs with a line of best fit. In addition, he also correctly placed measurement units outside the body of his results table and used single units (centimetres for length and seconds for time). In the last section he produced acceptable interpretations and conclusions about his data and made a good attempt to improve upon his evaluation. He recognised the group's difficulties with the equipment and finishing within the allocated time, but did not identify or acknowledge that these difficulties may have been caused in part by the lack of effective teamwork: “The problem with this experiment was time. We ran short of time, which made us more hurried and less careful. Because of this the experiment was slightly less accurate”. Martyn did not mention the ‘approximations’ they had made with their data due to running out of time. In attempting to link the observed trends in the data to the science ideas behind the pendulum, he was unable to use the science concepts to explain why the pendulum length affects the pendulum period. He instead tried unsuccessfully to explain why the pendulum swings in an arc. For this final part of the report he received a merit grade, and achievement with merit for the overall standard.

Peter was awarded achievement with merit for his summative grade, and he showed evidence of learning in all three sections of his report. His very detailed step-by-step method, which now included a list of equipment and the averaging of results, earned him excellence for first section. In his results section Peter failed to meet several criteria necessary for merit, including mention of the independent and dependent variables in his graph title and the line of best fit - he had attempted a curve, which was not the best fit for his data. However, because he had met so many of the excellence criteria, such as repetition (6 repeats for each pendulum length), and the use of average results Jenny took a holistic approach to marking this part and awarded an ‘on balance’ grade of merit. Teachers in the department had been given the discretion to use their judgement in this manner for situations like this. Peter’s results table now showed units outside the table, which was an improvement on his formative exercise. His interpretations and conclusion statements demonstrated a much deeper appreciation of what was required in this section of the report. He made appropriate interpretations based on his data and had partial success explaining his results in terms of the background science, including the roles of gravity and kinetic energy, and the relationship between the length of the pendulum, distance travelled and the period length: "Science ideas that explain the trend in the results link back to the force of gravity, from the horizontal position, gravity pulls the bob down. The shorter the string, the faster time it will take because it has less travelling distance". However, his next phrase shows an alternative conception when he goes on to explain that the bob on the shorter string "picks up speed faster which is why the shorter the string the quicker the period time". Peter also evaluated their experimental work, and identified appropriate amendments to the method, like changing to a "spherical bob", but like Martyn he did not divulge the group's fabrication of results. He was awarded excellence in this section.

Mitchell gained an overall achievement of merit. His plan was judged feasible, a similar result to his formative exercise. He had omitted to mention the averaging of multiple pendulum swings to obtain the period of one swing and timing one run at each pendulum length and then repeating the whole exercise six times was theoretically correct, but impracticable. Like Peter, Mitchell missed some merit criteria in his results (for example, drawing a curve line instead of a straight line on his graph), but because he had met so many excellence criteria Jenny awarded him merit. He too earned excellence in his final section, clearly identifying the group's problem with the efficient working of the pendulum and suggesting appropriate means of improving the situation (the use of a longer rod at the top of the clamp stand to prevent the pendulum swinging into the stand). He did not mention the results they had fabricated. His attempt to link the observed trend to the background science showed he had some understanding of the principles involved: “The reason for the trend in our results, that the time of the pendulum to swing increases as the length of the string increases, is that the long piece of string has longer to travel and loses its gravitational energy much quicker”.

Despite the difficulties I observed with Eddie during the performance of this task, he was able to achieve the standard at ‘achievement’ level. In his planning section he was able to correctly identify the independent and dependent variables and he gave sufficient detail for his plan to be judged feasible (merit level). A diagram he included in this method illustrated his misunderstanding of the concept of period, which he depicted as only the one-way swing of the pendulum rather than the return trip as well. In his formative report he had not drawn up a table for his results, but he had in this write-up, even including a column for averaging. His graphing work, however, showed a number of incorrect features. For example, he had used a bar graph instead of a line graph and his heading had no mention of the variables being investigated. Jenny’s awarding of an achieved grade for this part of Eddie’s report appeared to be another example of holistic marking since his account did meet many merit and excellence criteria, such as inclusion of relevant data and averaging. In the last section Eddie was able to provide correct statements for both the interpretation and conclusion parts, and he mentioned the pendulum difficulties the group experienced. His assertion that there were no systematic errors suggests he did not fully comprehend this concept, since he did not make the link with the problematic pivoting action of the pendulum and its potential to introduce errors. While there were no links between the findings and the background science, he did suggest doing fewer repeats of each swing (three instead of six): “Then if you did it only 3-4 times you would still get a reasonably accurate answer and save heaps of time". Jenny judged that on balance Eddie had achieved at merit standard for this discussion part.

Sam succeeded in the standard at merit level overall. His planning and method description was very long and detailed and very close to excellence level. Lack of precision in his English communication detracted from the quality of his report at times, but the only point missed in his account of the experimental method was the requirement for the pendulum to be free-swinging at its pivot point. His results section scored excellence despite some slightly confusing statements. This appeared to be an instance where Jenny gives the student the benefit, judging him to have met the criteria. The title for his graph was at first glance nonsensical: "The length of the pendulum (cm) to take one period in (seconds)", but he had mentioned the two key variables (pendulum length and period time). In his conclusion section Sam's inability to express himself clearly at times definitely hindered his achievement. His conclusion implied correct understanding but it was awkwardly expressed. Again in his discussion section he supplied appropriate pieces of information, such as means of addressing the knot tying difficulties they experienced and explanations for the observed relationship between pendulum and period lengths, but not in a coherent manner. He was given 'achievement with merit' by Jenny.

When I talked to the students later about the summative assessment the feeling amongst the group was that the practical session did not go well: "Like we were stressing hard, we ran out of time". Next time Martyn said he would "spend less time messing around". Eddie attributed part of the time wastage to the excessive number of repeats done in their experimental work, so for each pendulum length instead of performing six runs he recommended three. Sam had actually missed doing the experiment with his group because he was absent that period. He found not having experienced the practical aspects of the investigation hampered his ability to do the write-up well: “You can get all the information, but still you don’t actually know what even happened to the bob … I would have liked to have done that experiment”.

4.2.4.11 The reflections of the director

In her last interview Jenny was very pleased with the overall performance of her students. Their ability to write up reports particularly impressed her: “Their methods have improved incredibly and they now understand how much detail you need to have … most still didn’t get excellence, but when you read them, they really are superb … they’re almost to an excellence”. Jenny was also very gratified to see students, not only identifying errors in their evaluation sections but also coming up with “really nice solutions … they’re actually thinking about what I would do better next time I do this particular experiment, which is great because in reality as scientists that’s what you do”. She acknowledged the influence the nature of the background science concepts had on students’ ability to link their findings to existing scientific theory. In her experience students found using the collision theory to explain findings in the formative assessment a much more straight-forward task than trying to make links with the theory of the pendulum in the summative assessment: “A lot of them found the science just difficult to explain”.

In Jenny’s opinion Peter, Mitchell and Sam showed great “maturity” in their approach to the assessment – they were focused and “tried really, really hard”. In class these three students “are on task, they are listening, they are participating in class discussions, they ask me questions and they actually will do their homework as well". Jenny sensed their attentiveness in class indicated these students wanted to learn and succeed in their learning. Jenny attributed a lack of attention to detail for Martyn's failure to achieve higher grades, and insufficient preparation for the assessment on Eddie' part.

Jenny was divided in her opinion about the value of the peer assessment component of the formative assessment as a teaching and learning strategy. In hindsight she felt it would have been wiser to mark the students' work herself, and write her comments on their papers to avoid the common mistakes, such as not ruling a line of best fit on their graphs, students were making: "Like that graph error, I would have circled it … I've never had students all do that graph error". She would also have gone through the marking with the whole class using an overhead of the assessment schedule. She was concerned that very few students had approached her after the peer assessment, despite her invitation, to check or query their marking.

Formative-summative tensions really only arose for Jenny when it came to her role during the summative assessment. The departmental guidelines and discussions had given her clear indications about most aspects of this role, including the pre-teaching of background science concepts and support for students in preparing them for the final assessment sessions. In the summative sessions she was to act as an examiner, and she admitted she found the boundary between 'coach' and 'referee' a little difficult at times. When students asked questions she did her best to act professionally, but there were instances when she felt it was appropriate to help: "One of my boys didn't understand a particular word because English is not his first language", so she helped with his reading to determine the word's meaning.

4.2.4.12 The actors' impressions

In the third and last interview, after the summative assessment, Sam joined the discussion with Martyn, Peter, Mitchell, and Eddie. When we talked about any further learning occurring since the last interview, the group tended to reiterate earlier comments about how to write up the experiments and the order in which to do it. Martyn felt he now knew how to do the final sections of the write-up, while Peter thought he finally understood "what all the variables mean" and "which each one was". Peter also acknowledged that he recognised what kind of graph to use in his write-ups now, while Sam had learnt about "the different equipment that we can use", and "the different ways of using them". I probed further, but in terms of new learning the students had difficulty identifying new knowledge and skills. It appeared that in the students' minds they had made only small learning gains since the formative assessment.

While looking through the students' previous transcripts of interviews and observational notes of classroom activities, I had begun identifying strategies and techniques, and circumstances under which learning might be helped or hindered. I identified these items from information volunteered by the students and teacher in their interviews, teaching and learning materials, and my observations of classroom sessions. To add to my data richness, and confirm some of my tentative findings with my research participants, I decided to devise a questionnaire (see Appendix I) based on these items. I introduced the questionnaire to the students in this last interview to help gauge the extent to which the group agreed or disagreed about the efficacy of each item in assisting learning. The collated opinions of Martyn, Peter, Mitchell and Eddie are shown in Table 2 below. Sam was present at the interview but did not hand in his questionnaire.

The use of this instrument in the interview provided a focus for student discussion, as they sought clarification and verification of the meaning of certain items from each other and myself. The group's interactions, as they filled in the questionnaire, provided some valuable insights into their thinking. For example, having a knowledgeable teacher was generally considered "a good thing". Martyn thought that a teacher who knew what they were talking about was important because if a teacher was not knowledgeable "how do I know they are telling me the right thing? You just don't know whether to believe someone if they are not very good". A technique of questioning teachers use, known as "spotting", also generated discussion. Mitchell described the technique: "You could be just sitting there and then they just ask you". The group considered 'spotting' particularly effective for helping learning because it forces a student to concentrate in class. Peter felt it made you realise: "I don't know it, so you concentrate and you learn that".

Table 2
Item Efficacy in Assisting Learning

Item	Number of students who found the item			
	very helpful	helpful	sometimes helpful	unhelpful
Pre-tests	4	-	-	-
Knowing beforehand what you have to know	3	1	-	-
Things you already know	-	1	3	-
Learning about things that interest you	2	2	-	-
A knowledgeable teacher	2	2	-	-
A teacher you like and trust	2	2	-	-
A teacher who can control the class	2	1	1	-
Listening to the teacher explain	1	2	1	-
Understanding what the teacher says	3	-	1	-
The teacher asking you questions	2	1	1	-
The teacher answering questions	2	-	2	
The teacher giving notes	2	2	-	-
Lists of definitions	2	1	1	-
Teacher demonstrations	3	1	-	-
Tips from the teacher	2	2	-	-
Diagrams	2	2	-	-
Brainstorms	1	2	1	-
Mind maps	1	1	2	-
Summaries	3	1	-	-
Cheat sheets	4	-	-	-
Doing experiments	4	-	-	-
Doing practice assessments	3	1	-	-
Being taught the skills needed first	4	-	-	-
Being taught the background science	2	2	-	-
The planning template	3	1	-	-
Feedback from the teacher after marking	2	2	-	-
Marking other students' assessments	-	2	2	-
Assessment schedules	3	-	1	-
Giving feedback to other students	2	1	1	-
Feedback from other students	3	1	-	-
Working in groups	3	1	-	-
Working in pairs	4	-	-	-
Working alone	2	1	-	1
The student workbook*	2	1	-	-
Homework	-	2	-	2
Family	-	1	2	1
Post tests	1	1	2	-
Revision	1	2	2	-
Getting credits for NCEA	2	1	-	1
Getting merit and excellence for NCEA	3	-	1	-

* There was a non-response from one student on this item

The strategy of peer assessing was not identified as being particularly helpful for learning. Sam sometimes found it more confusing than helpful "because we're not sure having to mark" and as a result "you think you are all confused". He didn't think it was the students' role to mark: "We just go to school to work". Mitchell expressed similar views about the confusing nature of the marking for him: "Well, how can they get that right and I get that wrong and it's kind of the same?" Sam concluded "that's why science is not as good as maths … like maths has only got one answer and science has got heaps". The value of homework caused some debate. Peter saw pros and cons – he found homework helpful if you knew what to do: "It's like a revision of the day", but agreed with Martyn that it was a "waste of time if you don't know what to do", and probably made things worse because the work done was likely to be incorrect.

All agreed working together was helpful if the group was on task. Martyn saw advantages because if you were unsure about something you could ask a fellow group member for help: "If you don't understand it, it's easier and it clears it up real well sometimes". There was a general consensus, however, that there was a downside to group work. Peter commented; "Sometimes it's good as, but sometimes you just muck around". Sam added that group work was not helpful for learning "if you're in a group you don't like". Again Peter considered working alone was good if, like his homework comment, "you knew what you were doing". Mathew summed up the group's view about working together or alone in his comment: "I reckon if it's in pairs with people you like then it's mostly better".

4.2.4.13 A 'successful' production

Jenny's team of actors turned on a performance good enough on the big night to satisfy the critics, despite a potentially disastrous case of stage fright. Although the actors were unsure of some moves and fumbled their lines in places, Jenny's careful preparatory work with them and direction in the lead up to the performance paid off. The drilling of lines and frequent rehearsals gave the actors the ability to 'ad lib', when the need arose, in such a fashion that they managed a more than passable overall performance.

4.3 The 'Mountain View High School' Expedition

4.3.1 The school

Mountain View High School (a pseudonym) is also a large, high decile (9) urban school, but the make-up of the student population differs markedly from that of the school in the first case study. In this co-educational secondary school, girls comprised 50% of the student body, and the three largest ethnic groups were New Zealand European (68%), Asian (15%) and Māori (12%). The prospectus identifies teaching and learning as the cornerstone of the school with high expectations for student learning and good behaviour. A recent Education Review Office (ERO) report comments that a positive learning culture pervades the school, and many students achieve highly at regional and national levels. Mountain View High School was also into its third year of implementation of the NCEA qualification at the time of my study.

In Year 11 all students were given the opportunity to opt into six subjects for NCEA, but their entry was conditional on teachers' confirmation that these students were academically able to handle the extra workload. Science was not a compulsory subject, but the majority of students in Year 11, 10 classes in total, took it up. The school also offered courses in biology and horticulture, and all sciences were assessed using achievement standards. Most students in Year 11 opted to take five subjects for NCEA, and the Year 11 science class selected for this case study came from this group of students. There was a wide range of abilities amongst students in the class, but all were considered capable of meeting the academic demands of a course based on achievement standards. The Level 1 course gave students access to a total of 29 credits towards NCEA, three credits more than the national norm for a Year 11 course. Some 23 credits came from six externally assessed achievement standards and the other 6 credits from two internally assessed standards. Included in the 23 credits for external standards was an additional chemistry standard compared to other schools. The ethnic mix of the class closely reflected that of the whole school population. There were several amongst the Asian students in the class, including one in my study group, for whom English was a second language.

The students experienced all their science lessons in an older style laboratory, but they sat in groups at tables, which were movable. The teacher placed herself at a bench between the students and the whiteboard at the front of the classroom. She did much of her instructional teaching from this position, and frequently moved around the classroom when students were engaged in practical work or set tasks. Amenities such as workbenches for experiments, and sinks and gas taps were located around the perimeter of the room. Common laboratory equipment and chemicals were accessible to students, in labelled side cupboards or on shelves near the front of the room. Student work and posters were displayed on two walls of the room. The laboratory opened out onto a grassed area surrounding a garden used for horticultural teaching and learning.

4.3.2 The teacher

Kathy (a pseudonym), the classroom teacher, had begun her teaching career three years earlier at Mountain View High School after completing a conjoint degree in science and teaching. She held an assistant teacher's position in the science department at the school, and her senior specialist subject was chemistry. Kathy had taught Level 1 NCEA Science for two years prior to this study, and felt very fortunate to have had the close guidance of her head of department during these first two years of implementation. On the basis of recommendations from moderators after the first year "we changed a few things but it stayed essentially the same" in the second year. She was confident that the senior teacher with new responsibility for overseeing Year 11 science assessment for NCEA would retain similar procedures and contexts for assessing achievement standard 1.1 *Carrying out a practical investigation with direction* in this third year.

When I asked Kathy in our first interview about her views on the teaching and learning of science, her own teaching and interactions with students took precedence. A strong sense of vocation was evident in many of her comments, and encapsulated in her observation that teaching was not a job for anyone: "I truly think, and I'm probably taking a risk here, that you are born a teacher. I think you know on your first placement whether you're going to like it or not and I loved it". She focused on her own goals in teaching, and the importance of her role as teacher: "I try my very best to be enthusiastic and positive, and promote that when they come into here this is now the time for learning ... they have to concentrate and pay attention". Kathy felt lucky to be teaching: "I really, really, wake up every morning and want to come to school", and concluded, "I am obviously doing something right because the students respond".

In articulating her views of teaching Kathy did not volunteer many opinions about the nature of science education or science, but she did view science as a context that offered opportunities for the use of teaching strategies that resulted in effective student learning. For example, in her teaching approach she valued the hands-on learning that science experiences could offer: "I try and use lots of different techniques and activities with them to help students. We are lucky in science where we can do heaps of hands on … I am doing practical work all the time for that reason because most of them seem to learn better that way". When I asked her late in the interview what she believed good investigatory science to be, she considered this for some time before replying; "Well I guess it is all about coming up with a question and trying to prove or disprove it".

Kathy stressed the need for lots of different activities to cater for students at all levels as a key consideration in her teaching approach: "We've got to teach not only to the bright but to the ones that struggle along, and I try and help them a lot". In this regard she recognised the value of students working in groups. During group work early in the year if Kathy noticed quieter students who tended to "leave it up to the others to do", she made sure that she had a say in future decisions about the compositions of groups. She ensured that these reticent students "were going to be with people they thought they would be confident enough with, but also people that would guide them if they were perhaps a bit limited and help them in that way … you get involved and you learn from your peers". She justified her intervention in these decisions, otherwise "you just get students that would be left behind". Kathy also acknowledged the presence of students from many cultures and their needs: "I always try and think up different analogies to try and help those who don't or can't relate to that experience that I have experienced". She judged the effectiveness of her teaching by "the marks you get at the end", although there were other indicators: "If I was having huge numbers of questions and evidence that there was confusion, like with body language and just misbehaviour or if they weren't understanding, then I would have to think that I was doing something wrong, that they just weren't getting it". In such situations she analysed her teaching: "It was too hard or I wasn't able to explain it right or they weren't listening". As a result of such experiences, Kathy felt she had learned the necessity of being flexible in her teaching. She recalled a fourth form class from the previous year who, unlike her other classes, was not responding well to her teacher directed approach: "They just didn't want to know. So you've got to be adaptable … I thought this isn't working and I changed everything". The changes proved successful and the students took an active interest: "We ended up making videos to teach weather, they did weather reports".

Kathy placed great store on creating a relaxed and happy learning environment for students "where they can feel comfortable and safe and where they can actually learn". To achieve this end Kathy identified humour, praise, and mutual respect and trust as important components of her teaching: "Humour is a big thing, heaps of praise, but also firm when you need to be, and also respecting their ideas and their values as well. Trust is big". She learned the value of praise in promoting learning from her observations of other teachers in action: "I praise a lot and I tell students if they are doing well", and as a result "I think you get respect". As well as "showing them that you know what you are talking about and also that that you have passion for your subject". Kathy maintained you had to demonstrate "boundaries and guidelines, and be fair and consistent all the time" to earn their respect. She advocated strongly the teacher's role in promoting appropriate student behaviours and manners: "I constantly remind students to use their manners because that is important". These qualities and actions were part of what Kathy termed being "very professional" as a teacher. Kathy also believed the physical environment played an important part in creation of a classroom atmosphere conducive to learning. She used displays on topics of interest and relevance to students, such as sport, and samples of students' own work and plants to create such an atmosphere.

Kathy recognised that the school culture had an influence on her philosophy of teaching and learning: "We're lucky because we are in a good school … we are very academic driven so there is the expectation as a teacher at the school that I push that. I have to have high expectations – make sure homework is done, books are up to standard". School structures and systems like the Dean System nurture this academic culture. The Dean System was composed of designated teaching staff responsible for student welfare at each form level in the school. Kathy commented; "It's an awesome Dean System, and being responsible for a form class I'll have teachers come up to me, that are teaching these students, and say 'so and so' hasn't done their homework … or was very rude to me today". By following up on such student behaviours in her form teacher's role Kathy accepted the school's philosophical stance on academic achievement, and her responsibility in helping students to meet the school's expectations. She was conscientious in this role because she sensed in the school "there are huge expectations and philosophies that we follow" regarding students' behaviours.

Assessment, in Kathy's view, was an aspect of teaching which she felt obliged to do: "I mean I have to do it, we have to report to parents. The students have to know how they are doing - we need to know. We test them on every unit they do, and we mark, we give it back to them, they file it away and we report it to parents". She appeared not to attribute assessment with a role in assisting learning, but rather with reporting achievement. However, in her comments about summative testing, she did identify a formative use of the assessment information, acknowledging she used this data to assist some students with their learning in tutorials: "As I mark tests, I look at them, I talk to a student about it if I think they've done poorly – why haven't you done so well? Did you study? Is there anything I need to help you with? Why don't you come to the (Tuesday) tutorial and I will give you some extra help". Kathy in fact judged the effectiveness of her teaching ultimately on the marks students obtained in their assessments, although she recognised "there are lots of factors as to why students don't achieve, it's not just the teacher".

When we talked specifically about assessment of the internal Achievement Standard 1.1, Kathy began describing more assessment practices that had quite clear formative functions. For example, in previous years she had monitored students' learning progress by looking at work they had done in their laboratory manuals. This laboratory manual work included a planning exercise for a practical investigation; "I actually look through it and write on it, and tell them what they need to do". Also, all students in Year 11 Science at the school do a mock assessment known as the 'formative assessment', which is done before the summative assessment event. The classroom teachers, including Kathy, mark these investigations and use the results to inform some teaching decisions: "We look at the marks and use them to help us see who they (the students) should be working with (for the summative assessment event), or if they are going to need some extra guidance or direction". Departmental policy was to place students, on the basis of results in the formative exercises, in groups of mixed ability for summative assessments to maximise all students' chances of success in this internally assessed standard.

When I asked Kathy in our second interview how she knew if students were learning she clearly articulated formative assessment strategies, she said: "I will ask them different questions as I walk around the classroom and wait for them to ask me direct questions and if I get the correct answer then I am happy that they have learnt something". She revealed that the mock assessment provided her with "very valuable" information about gaps in the students' procedural knowledge, like not "clarifying the variables and how they were going to control them to make them a fair test". On the basis of this information Kathy monitored their understanding and awareness of correct procedures by using opportunities when students were working to "actually physically go right around ask every single one of them … just probing them all the time and asking questions again and again to see if they are actually learning it". If she found patterns in their lack of understanding, or "I am noticing they are confused in a particular way" she stopped the class and did some instructional teaching by way of explaining and clarifying. She thought it important to remind them all the time of what was required of them: “I use the word hint … you must remember to do this and if that’s the instruction then you need to make a statement”. She also wrote comments on their formative assessment script “so they can then take them away and read it on their own and reflect on it as well”.

One consequence of the NCEA system, which Kathy was unsure about in her initial interview, was the high priority students appeared to place on simply gathering credits rather than striving for higher grades; “I don’t know how good or bad that is. I don’t know”. Later, as teaching of the topic progressed, this attitude of students towards NCEA proved to be an area of concern for her (see Section 4.3.4.6).

Kathy’s beliefs about teaching and learning in science focused on those aspects of teaching practice that she believed impacted positively on the quality of the education students received, and which she sought to emulate in her own role as a science teacher. The vocational aspect of teaching, as a calling with responsibilities that included the teaching of social values, came through strongly as a core belief for Kathy. The nurturing of positive teacher-student relationships and interactions, and support for students in achieving learning goals such as attainment of qualifications, played pivotal roles in her perception of successful teaching and learning.

4.3.3 The students

After my introduction and talk to the class about my research project, Kathy followed up my invitation to join the research group with a personal approach to a number of students she thought would be amenable and willing to share their learning experiences. As a result four students came forward as volunteers. These original group members included two female students – Anne, a New Zealand European and Carol, a recently emigrated Asian student for whom English was a second language (often referred to as ESOL students). The two male students were Alex, an Asian student educated in Australia and New Zealand with well-developed English language capabilities and Mark, a New Zealand European. Three members of the group (Anne, Carol and Alex) remained participants for the duration of the study but Mark withdrew after the first week. Steve, another New Zealand European male, took his place.

As it transpired, for a number of reasons I was unable to obtain the rich data about the day-to-day learning experiences and thoughts of my participants, and the continuity of data that I had hoped for in this case study. These reasons included:

- the rescheduling of the teaching timetable for the unit which prevented me observing most of the formative assessment exercise. The initial set of 4-5 lessons, devoted to the teaching of the content and the three lessons for the formative assessment, were originally scheduled for late in the first term. The summative assessment was to occur approximately eight weeks later, during the school's internal exams period in the second term. Just before teaching of the unit was due to commence the department decided to shift the unit into Term 2 to allow the formative assessment to be closer to the summative event. Unfortunately the work arrangements I had made, to allow me to be present during the teaching and assessment period, could not be altered. The problem was partially solved by the generosity of Kathy, who went ahead with the teaching of the content segment of the unit, after seeking the permission of the head of department. The formative assessment, however, had to be performed at the rescheduled time in line with other classes. Consequently, while I was able to witness the instructional teaching sessions, I missed first-hand observations of students working during most of the formative assessments

- my inability to observe my four participants working together as a group. I was only able to view Anne and Christine working together during some of the instructional sessions because Anne was absent from class for several periods participating in the school drama production. Alex, Mark and Steve all worked with other members of the class, and in separate groups. For the assessed practical investigations (formative and summative) Kathy placed my participant students in different groups, on the basis of their performance in the planning component of the activities, so it was only possible for me at any given time to make close observations of one student. This focus on individual students also restricted my ability to observe and record observations of my participant students carrying out similar parts of their investigations. For example, during the data-gathering phase of the formative assessment I could only focus on Carol's actions – the logistics of the room and experimental set-up for the assessment made it difficult for me to move from group to group, so I chose to concentrate on Carol's group. Thus I was unable to collect observational data related to the performance of experimental work by my other participant students.
- the school's ruling that outside observers could not be present during summative assessments. This eliminated me from first-hand experience of the students' actions and behaviours during the final assessment.

The rather disjointed nature of my observations left me feeling insecure about the validity of my data, because I felt distant from the students and their experiences. It was at this point that the value of using triangulation of data gathering methods and sources became evident, because my other data sources helped to fill some of the gaps in my understanding of student experiences. In one lesson taken by a relief teacher, I took the opportunity to chat with individual members of my group to extract more information from them about their experiences. These informal interviews provided valuable data. In lieu of my presence in certain classes I designed a simple one-page questionnaires for my participants to fill in covering aspects of their formative experience I could not directly observe – two students were able to contribute data in this way. I also arranged for Kathy to audiotape a lesson for me, and followed up the lesson with a telephone conversation between Kathy and myself.

The first formal interview took place with the students half way through Term 1. All four students seemed a little apprehensive and diffident during the interview, especially when I asked open-ended questions of the whole group. I was able to overcome their hesitancy to a certain extent, by directing questions at individuals and rephrasing the questions to gently probe and prompt. Anne and Mark became more comfortable and confident in their remarks, and Alex did respond if directly asked. Carol's limited spoken English clearly affected her ability to express thoughts fully, but she did manage to convey some valuable insights into her experiences learning in the science classroom.

Later in class, I observed that Anne and Alex went about classroom learning activities confidently and experienced few difficulties. Alex was a very capable student in science. In his interview he expressed his enjoyment of science: “It’s quite fun, like you do practical stuff as well as theories”. He considered that he learned well in science; “It’s interesting and it’s easy. I just seem to go and grasp the concepts easily … it doesn’t take long to understand”. He intended to do a degree in sports science because he wanted a career as a sports journalist. In class Alex worked with another capable student, and together they frequently successfully completed tasks ahead of others in the class. He valued being able to talk things through with his partner, especially if he required further explanation. Alex found that “if you don’t really quite understand some terms” his partner could explain them in another way and “you can see his point of view sometimes”. Alex operated largely independently of Kathy, his teacher, and his interactions with her in the whole class forum were minimal.

Anne was finding science easy because “last year I had a really good teacher and so I learnt heaps and now it just comes easy … we were into some Fifth Form stuff last year, it’s kind of easy this year”. In class Anne was an outgoing student and I frequently observed her in the early lessons conversing socially with other students while the class was doing set exercises. Despite instances of off task behaviour, and later absences from class due to school commitments (the school’s drama production), Anne managed her workload well and achieved highly in assessments. She had periods in class where she worked intently. If she had difficulties with tasks she actively sought out assistance from peers or the teacher, and she was often active in class discussion work.

Carol said she enjoyed science and needed to learn science because she wanted to be an industrial designer. In class she actively involved herself in activities, but rarely participated in whole class discussions. In practical work she followed the lead of other students, and in written tasks persevered for long periods of time, often referring to an electronic dictionary that she had with her at all times. I noticed her in one lesson seeking help from Anne with the meaning of particular words and phrases, and Anne willingly obliged. When I asked her how well she learned in science Carol was not confident about her learning, but recognised that “my problem is language, so I don’t understand some of it sometimes”. If Kathy approached her in class, she would take the opportunity to ask questions or seek assistance, but again not in the whole class forum.

Mark said he did not enjoy a lot of learning success in science because “I have trouble paying attention sometimes I just can’t concentrate. I’m not very smart … I just don’t find some things interesting. I like to do more extreme stuff, hands on kind of person … I just get bored and have a chat”. He was not intending to do anything later in life in science, thinking he might do an apprenticeship. His behaviour in class reflected his comments about needing to be ‘hands on’ – in practical activities he was animated and involved, but in tasks requiring him to read and/or write he appeared unable to concentrate. In these situations he was either easily distracted, or became a source of distraction himself for other students. He would often seek Kathy’s attention calling out questions like “Where are the glasses Miss?”, and seeking confirmation over procedures. In the second lesson he was removed from class for inappropriate behaviour. It was after this lesson that Kathy informed me Mark was withdrawing from the study because his family was taking him overseas, and he was going to miss both the formative and summative assessments. Kathy had arranged for another student, Steve, to join the study.

Steve was a student who frequently sought the assistance of his peers in carrying out learning tasks. He appeared to experience difficulty with aspects of the work, but often worked productively with his neighbour to complete set exercises. He rarely sought Kathy's assistance, but did respond freely when Kathy approached him during her movement around the room during classroom sessions. During activities set from the textbook he had trouble staying focused in class, and I observed him chatting socially to other students. However, he was rarely the source of distraction.

When we moved on in the first interview to discuss the type of teaching approach Kathy used, Anne was able to describe in some detail the manner in which Kathy usually introduced topics since Kathy had also taught Anne when she was in Year 9: "She generally does an overview of the topic and then goes into extra detail – gives us a bit of a summary (orally)". For the current topic, their teacher also "showed us the objectives or what we should be learning in our lab manual".

Mark's comments about Kathy's teaching approach revolved around his thoughts of her as a teacher: "I reckon she's pretty good … she explains things well, and she doesn't yell at you if you do something wrong". Anne agreed; "She really understands. If you don't get something she writes it down and explains it until you get it". Anne also believed she was a much better teacher now than she was back in Year 9: "She was a first year teacher then, so she didn't have many teaching strategies. She wasn't too great, but now she's a third year teacher she's kind of got into the habit, so she's pretty good. She was too nice in the Third Form. I'd just sit there and read a book, but now it's kind of like you've got to do the work". Alex's opinion was that "she's not so aggressive, like some teachers".

At this point in the discussion I asked the group what helped them to learn, and each student had clear ideas about what helped them personally to learn. They all agreed the material being learned needed to be interesting, and all enjoyed experimental work. Mark needed hands on work, while Anne found variety in classroom strategies and activities helped her learning: "Vary between writing and like doing exercises. If you have to write for pages and pages you just lose interest and you stop reading what you are learning. But if you do exercises and you kinda put it into practice then you're learning and then you do a bit more writing". Alex did not favour too much writing either, finding it "kinda repetitive". Vocabulary lists and "some examples of it [full sentences containing the word in context]" were especially valuable for Carol. When Alex said he preferred classroom environments where some chatter was going on, but not overly noisy because it was hard to work, the entire group agreed with him. Anne found completely silent classroom environments off putting: "When everyone is silent it's like really weird you know", and she went on to comment how awkward you felt about asking for help because everyone could hear. Classmates generally helped my group members learn, but for Mark they were often a hindrance: "If I'm bored we just talk for ages and get behind in our work".

For three students, their families played a significant role in promoting their science learning. Mark's mother was a scientist and "she helps me with my science … and she stresses me for my exams. I actually study". Carol volunteered that "my dad is helpful, because he's a chemistry teacher, so he helps me", and she asked him "lots of questions". Alex's younger brother was one year behind him at school "so sometimes he asks me about these questions. They're kind of revision for me". Anne explained that her parents don't really understand science, and while her stepmother was very helpful; "I grasp the concepts quite quickly".

The students appreciated particular teaching strategies that Kathy used like bingo, matching games, and special little tricks for making learning easier. Anne found a certain method for writing chemical formula that Kathy used "quite helpful. You learn how to write them properly". Alex commented in passing that when Kathy linked the science ideas to "everyday things" that helped him learn, and the others agreed. Summaries that Kathy sometimes provided at the end of a topic before a test were helpful according to Anne because "you kinda know what you have to know for your test".

Finally we discussed NCEA, and comments centred on the qualification's system of awarding achievement. The students had mixed feelings. Mark didn't fully understand how grades were obtained, but thought it may be better than School Certificate where "if you get 49% and you don't get over 50% you don't pass and you've really bombed out". He thought NCEA had some system whereby they "minus merits and put it onto 'achievement'", and as a consequence you were less likely to fail. Alex preferred "to have a percentage total, it's easier for me to know how I've done … because I'm not too familiar with it [NCEA]". Anne recounted that "last year everyone was really nervous about going over to NCEA … but now we're kind of in it, and we've had a couple of practice tests it's not too bad, because you learn as you go. So no, it's not too bad". Carol favoured what she termed the "percentage system" because in NCEA "you can't learn how well you do". She worried that even if you obtained an excellence you could not tell if it was a "high or low excellence".

4.3.4 The story as it unfolds

I have likened this story to a climbing expedition, where Kathy is a motivated and enthusiastic mountain guide, committed to helping young climbers achieve success in their mountaineering endeavours. In this story the mountain to climb is Science Achievement Standard 1.1. *Carrying out a practical investigation with direction*, and Kathy's students are a group of young climbers with a range of abilities and experience. Kathy's primary goal is to ensure that everyone reaches base camp (secures credits for the standard), through students working as teams under her guidance. She monitors individual progress carefully during practice climbs, and groups students for the final ascent so each climbing team has a highly capable climber leading by example. With increased confidence and skill, individual students are able to achieve even greater personal summits at 'merit and 'excellence' levels.

4.3.4.1 The mountain guide's plans and strategies

From my interviews with Kathy, observations of her teaching in class, and examination of documentation she made available to me, it was clear that a teaching sequence devised by the Head of Department (HOD) was a major influence on her classroom delivery for Science Achievement Standard 1.1. The HOD had taken the standard and broken it down into a set of four or five lessons. The lessons set in a chemistry context and covering concepts to do with rates of reaction utilised the following resource materials:

- a chapter from a textbook (Hannay et al., 2002) on scientific investigations, covering the aspects of focusing and planning, information gathering, interpreting results and reporting the investigation, and
- a section from a student workbook publication (Cooper, Hume & Abbott, 2002), featuring material focused on Science Achievement Standard 1.1 *Carrying out a practical investigation with direction.*

In her planning Kathy also:

- consulted guidelines from the teacher with responsibility for the Year 11 science programme for the internal assessment of this standard
- called on her previous teaching experience preparing students for this standard, and with Year 9 and 10 students doing science fair projects as part of the junior science programme at the school, and
- reflected on feedback from her classroom teaching, especially information she obtained from the formative assessment activity including students' written scripts.

Kathy emphasised that her planning focused on the Achievement Standard: "I basically teach it as a unit". In her words, she taught it "as a real assessment", equipping the students with the concepts and skills required for the assessment tasks. The decision to adopt this particular teaching approach was being influenced by demands being placed on programme delivery. The full teaching programme was one source of these demands: "We are only talking about four credits and we have got to remember that the externals are worth credits as well, so we can't spend any longer because we have got all the other topics as well". Another demand came from the students. Kathy was strongly of the opinion that if activities and assessments did not count towards NCEA credits, then students were not prepared to put in the effort to learn: "The students want to know for every test, every assignment you give them … is it worth credits? Is it worth credits?" While the standard was delivered as a unit of work, Kathy did point out that the science behind the science investigations, that is, the chemistry concepts and skills related to rates of reactions and specific reactions had already been taught in an earlier topic geared to the externally assessed Science Achievement Standard 1.4 (Chemistry).

Following the HOD's lesson sequence, Kathy relied heavily on the two teaching and learning resources to directly deliver her intended curriculum in class. The student workbook publication (Cooper, Hume & Abbott, 2002) was utilised initially to introduce and breakdown the standard's requirements into achievable steps for students. The authors of the workbook had adhered closely to the standard's achievement criteria and explanatory notes, and exemplar materials provided by NZQA such as assessment activities, schedules and templates. A list of specific learning outcomes the authors had derived from the standard served as a means of communicating what Kathy wanted students to learn – her intended curriculum. The workbook was also used as the source of many learning exercises, particularly practical activities. In class Kathy frequently consulted the teacher's version, which contained explanatory notes and answers. Text and activities were presented in a sequence designed to scaffold the learning of concepts, skills and procedural knowledge, which in the authors' interpretation met the requirements of this internal standard. In classroom lessons, and the setting of homework tasks, Kathy followed this sequence and selected most items.

The student textbook (Hannay et al., 2002), which Kathy used in class to reinforce her teaching and consolidate student learning, was written as a study guide for students. The first chapter was devoted to science investigations. In its format for this chapter, the text showed strong links to the SiNZC content, utilising headings directly from the Investigative Skills and Attitudes strand. The content systematically covered the four themes of focusing and planning, information gathering, processing and interpreting and reporting. Text content was complemented by the frequent use of exemplars to illustrate specific concepts and skills, and the inclusion of problem solving exercises for students to practise skills and apply knowledge. The emphasis was on fair testing, controls and experimental plans in the focusing and planning section. Experimental reliability, tabulation of data and the drawing and interpreting of line graphs featured in data processing, including the identification of cause-effect relationships. Reporting dealt with the generic aspects of stating an aim, hypothesis, method used, results and how data was processed and a conclusion and evaluation of the findings. Within the text there is no specific reference to the achievement standard, but in the introduction to the book readers are informed that this updated edition ensures full coverage and understanding of all the Level 1 Science Achievement Standards. A table also shows how each chapter links to particular standards. Kathy did not use this text as frequently as the student workbook, but did on occasion use it in whole class shared reading sessions to recap concepts. She did, however, set extensive work from it when she was absent from classes. Students were required to read sections and complete exercises in class and for homework.

Other departmental guidelines, apart from the HOD's lesson sequence, centred on instructions for carrying out of the formative and summative assessments. All teachers were to follow these instructions faithfully for moderation purposes. Both formative and summative assessments had similar contexts, where students were placed in a scenario situation as research scientists reporting back to a manufacturing company via a letter. The assessment tasks were modelled on exemplars provided by NZQA and mirrored the procedures used in the two previous years of NCEA implementation at the school. Teachers were to spend at least three lessons on the formative assessment in class, and the summative assessment was completed in a two hour time slot during the scheduled mid-year exam period. Kathy adhered closely to this timeframe. The formative assessment was marked by classroom teachers, while the summative assessment for the whole Year 11 science cohort in the school was marked by a panel of 3-4 teachers from the science department, using a common marking schedule.

Kathy used a mix of strategies in her teaching approach including:

- instructional sessions where she provided explanations, guided shared reading of text excerpts, led question and answer discussions, and fed back assessment information
- teacher directed activities where students were given opportunities for hands-on practical and planning skills development, or practice and application of knowledge through questions and exercises set from text, and
- one-on one facilitating of learning with individual students or small groups.

In her teaching Kathy kept students focused on task requirements, which meant in the workbook tasks because of their very nature, that students were concentrating on meeting the achievement criteria of the standard. These workbook tasks often contained templates for student responses and assessment schedules that were modelled directly on those provided for NCEA assessments, so students became practiced in their use. On the other hand the text exercises contained no overt reference to the standard's achievement criteria, only the principles of effective performance of investigations. Kathy did refer directly to the specific requirements required for the three levels of achievement in an assessment task, when she gave full and detailed feedback and feedforward to the class after the formative assessment. This information related to the requirements as spelled out in the formative assessment schedule, and formed the content of the last lesson before the summative assessment event.

In Kathy's mind, the purpose of the formative assessment was to provide students with an opportunity to experience summative assessment conditions, but under the careful direction of the teacher. It was to be a practice run for the summative assessment: "It's demoed … we go through the plan, we expose the students to the apparatus that they need and the chemicals, with labelled bottles … I hold it up, this is your thermometer, this is your 50mL beaker". The formative is marked "properly" by the teachers, to criteria similar to that used for the summative assessment: "We just monitor if they have made any errors, make sure they can change them before [the summative], and understand it before the real one".

The key concepts, skills and procedural knowledge that Jenny intended her students to learn are summarised in Table 3. This content was identified in data collected from her interviews, observation of classroom lessons, departmental teaching guidelines, and the student workbook (Cooper, Hume & Abbott, 2002) and textbook (Hannay et al., 2002).

4.3.4.2 Revising the basic rules of ascent

The early lessons went very much as Kathy intended, as she led them through the learning activities from the student workbook (Cooper, Hume & Abbott, 2002) and the textbook (Hannay et al., 2002). In the first lesson Kathy began the instruction by leading a shared reading of the introduction to the standard in the student workbook. As well as introducing them to the standard, and the requirements for achievement at the different levels, she also outlined the upcoming timetable for coverage of the topic including the timing of the formative and summative assessments. To elucidate terms like ‘feasible’ and ‘workable’ Kathy referred students back to their Year 10 science Fair projects, and asked questions like “what does process mean?” or, “analyse. What does that mean?” Student responses showed extensive prior knowledge of relevant concepts and skills. For example, when Kathy asked what could be done with data in the form of a set of percentages responses included “find the mean and median” and “rank them or graph them”. She encouraged students to underline or highlight key terms in the reading, such as ‘variables’ and ‘range’, and drew their attention to the list of specific learning outcomes that had been provided: “You need to look at these while you’re learning over the next year and tick them off as you learn”. A key message from Kathy was: “We’re aiming for excellence”. The shared reading was followed by a series of practical activities from the workbook, designed to practise important practical skills. The class were excited at the prospect of doing the experiments.

Table 3

The Teacher's Intended Learning (Case Study B)

Concepts	Skills	Procedural Knowledge
• Fair tests • Controls • Experimental plan which includes an aim, list of equipment and an experimental method • Purpose of an investigation as an aim, testable question, hypothesis or prediction • Variables - key, dependent and independent • Primary and secondary data, qualitative and quantitative data, reliability of data/a reliable experiment • Tables as a systematic format for recording data • Graph types (bar and line); graph components such as title, x (independent variable) and y (dependent variable) axes, units and values for axes, plotted points, and lines of best fit • Relationship between two quantities when change in one causes change in the other • Sources of error • Formats for scientific reports • Equipment names, types and purpose • Background/contextual science concepts pertinent to the investigation e.g. nature of the reaction between metals and acids, factors affecting rate of reaction	• Designing, evaluating, modifying and carrying out a systematic plan for a fair test • Determining the purpose of a fair test investigation • Identifying, controlling, changing, observing and measuring variables • Choosing and using equipment appropriately • Determining appropriate range of values for variables • Repeating experiments • Recording and processing data – tabulating, averaging, graphing • Interpreting data, and recognising trends and patterns • Discussing findings, linking findings to existing science ideas and drawing conclusions in a written report • Evaluating the investigation in the written report (sources of error, improvements)	• Knowing how to plan a workable, fair test • Knowing that planning requires trailing, evaluating and modifying • Knowing why reliable data is needed and how to obtain consistent data • Knowing that the findings should be linked to the purpose of the investigation and to the science behind the investigation • Knowing when assumptions can be made and the limitations of those assumptions • Knowing how to work as a team • Knowing how to interpret the template and assessment schedule requirements of tasks for the internal Science A.S. 1.1 at achievement, merit, and excellence levels

From the manner in which my study participants responded and behaved during these first teaching episodes I was able to detect some clear indicators of their levels of understanding and the learning strategies they used. Alex did not contribute much in whole class discussion, or appear to pay much overt attention to the material being read, but quickly got to work with his neighbour when Kathy directed them to the practical exercises. He and his partner accomplished these tasks with ease and interest, and sought little help from the teacher. Mark showed minimal engagement during the shared reading session, but became very animated and involved during the practicals. He worked enthusiastically at first with three other boys as they performed experiments burning magnesium, but later moved from this group to Alan's where stopwatches were being used. Throughout the activity he showed a keen interest and willingness to handle the materials and equipment, and moved amongst several groups. He sought out the teacher as she worked with groups, and asked her many questions often about equipment. Carol meanwhile followed the reading in the student workbook very closely, often referring to her electronic dictionary. In the practical activity she began working quietly on her own, diligently observing and recording the melting of ice, and later joined two other male Asian students in a dart making exercise. She did not make a dart, but was happy to participate with the boys in the test flights by timing and recording flight times. The trio spoke amongst themselves with ease in a Korean dialect. Anne was not present in this lesson, as she was involved in the school show rehearsal. Kathy drew the lesson to a close in a whole class forum, by calling for students' findings in each experiment, and providing extra information where required form the teacher's manual. Carol appeared to have significant gaps in her findings and jotted down many notes. Mark was often inattentive but did volunteer some answers to Kathy's questions.

The class took some settling in the second lesson after Kathy handed back the results of an electricity test they had done previously. In attempting to turn their full attention back to the investigation topic she commented: “You’re wasting a lot of time so far. This is your real chance to get credits”. Once the students are refocused Kathy reminded them that not only are they required to plan and carry out an investigation for the standard, they must also write a report: “Part of reporting requires tabulating and processing, drawing graphs, comparing data and finding averages and means”. She briefly covered these key points by asking questions, reminding and recalling instances when the students have used these skills. She then set exercises from the student workbook involving tabulating and processing (averaging and graphing) of data, and roamed around the room interacting with the students. Some of these interactions involved behaviour management to focus inattentive students back onto the task, and Mark was often at the centre of these incidents. He had arrived in class without his workbook and Kathy insisted he do his work on loose sheets of paper and paste or copy them into his workbook later. Eventually his off-task, and in one instance inappropriate behaviour, caused Kathy to remove him from the room and send him to the Dean’s office. He did not return. However, most of Kathy’s interactions with the students that period involved responses to student queries, checking of progress, affirming and encouraging students, providing correct or model answers, and pronouncing words for ESOL students. Periodically she stopped the class to re teach some aspect or emphasise a point.

Throughout the lesson most students had periods where they became disengaged from the tasks and chatted amongst themselves. In contrast, both Alex and Carol stayed focused for the majority of the time on the set work. Alex collaborated with his friend as they steadily made progress through the exercises. At no stage did they seek Kathy's assistance. Carol worked alone, with frequent use of her electronic dictionary.

4.3.4.3 Back from a break in training

The third lesson was after a two-week holiday break. Anne was back in class and had caught up on missed work by copying a fellow student's notes. Kathy reminded the class that the exams, and the summative assessment for Achievement Standard 1.1, were six weeks away. She pointed out the internal standard "is worth a lot (four credits). You need to work hard because it's not far off. Let's recap what it's all about". To refresh their memories Kathy used the first few pages of the relevant chapter in the students' textbook in a whole class guided reading session. She read and explained about typical components of scientific investigations while students followed the text. This reading exercise was carried on under some duress with a steady trickle of students arriving late or leaving on various errands, and whispered conversations between students – consequently not all students attended to the task.

Kathy then turned students' attention from the textbook to a planning exercise in the workbook. Using a template modelled on that used in NCEA summative exemplars, students were to prepare a plan to answer the question: "Which Bunsen flame, blue or yellow, heats water the fastest?" Her strategy was to lead the whole class through the planning exercise section by section. By explaining, questioning, probing and prompting she elicited responses from students, which she in turn confirmed as appropriate, corrects or questions further. In one exchange she used literacy skills to unravel the meaning of terms for students: “What are variables? What is a little word that’s in there? Yes, vary”. In this manner Kathy scaffolded the students’ comprehension of what information was required in the identification of variables section. In another instance she prompted students to recall previous experiences with various types of Bunsen burners, and the means by which air supply to the flame is controlled. In the latter part of this exercise Kathy changed her approach, instructing class members to continue on with the exercise themselves, while she moved around the room checking on progress. She spent most of this time assisting individual students or small groups with specific aspects of the task.

My four participants, who now included Steve (Mark was away travelling with his family), had been quietly focused on getting their written responses down in their workbooks during the first phase of the lesson. Kathy approached Anne very soon after the class began work on their own, and helped Anne to think through the section on the dependent variable with prompt questions like; “Will you need the initial temperature?” Anne later told me that this conversation with Kathy had helped her to understand what was needed and improved her answer. However, as Kathy moved on to spend several minutes with Alex giving him feedback about his written answers, Anne was very soon was off task, engaging in social interactions with several male classmates. Carol meanwhile had partially completed the sections on variables, but had bypassed them to begin listing the equipment needed. Steve was working with his neighbour to complete his responses.

Kathy returned to Anne shortly after their first conversation and began discussion on the section to do with obtaining reliable results. She reminded Anne of their science fair experiences with investigations in the junior school, and used an example of repetition of experiments to explain the meaning of the term ‘reliable’. At this point Kathy alerted the whole class to the marking schedule for the exercise, which was on a following page in their workbook, and asks them to mark their plan: “Do it with a partner. You could get your partner to mark it, or you could mark it, but tell your neighbour what changes you are making. Write in the correct answers”. She told them: “You do not need to write out the method in full. Put repeat steps 1-4 and add diagrams”. Kathy continued her discussion with Anne, modelling what she would be looking for in Anne’s answers to give credits.

Anne and Carol paired up to carry out the assessment, and Anne gave Carol very constructive feedback about aspects of her plan. For example, Anne’s commented about the purpose of the investigation enabled Carol to write in her purpose statement, which was missing from her plan because she had not been sure what to write. Steve and Alex, and their respective partners, were similarly deeply engaged in their assessments of the plans. After a little confusion over instructions, Anne verified with Kathy that a method was also to be written up in the lesson: “Yes, it’s part of planning. Everything needs to be written down”. Anne resorted to social chat again with her neighbours and Steve looked to his neighbours for support and direction in writing a method. Carol and Alex concentrated on their write-up. Kathy commented in passing to Anne that “a lot of detail is needed in planning”.

Despite Kathy's facilitating role, many class members found it difficult to stay focused on the method write-up, so Kathy returned to whole class teaching to help students finish their methods. She told them: "A plan must be detailed enough so even a non-scientist could follow instructions", and recommended the use of numbered steps like "One. Measure out 50mL of water using a measuring cylinder". Students were asked to come up with similar steps, and she gradually built on these responses until a series of suitable steps was attained. Completion of the method write-up was set for homework.

4.3.4.4 The mountain guide is called away

Kathy was unable to be present for lessons four and five due to a family funeral. Two relief teachers took the class instead, Mr Ashcroft (a pseudonym) for the first lesson and Mrs Smith (a pseudonym) for the second. Kathy had set work for the class, and Mr Ashcroft sought the details from Alan and his partner. It then appeared that not all students understood what work was to be done, so Mr Ashcroft gave them guidance starting the task. The work involved a planning exercise from the student workbook, where a problem was posed in a scenario setting, followed by processing and interpreting of given data and discussion of findings. The problem was to find a fabric type, which would make the warmest jacket, and the hypothesis being tested was that polar fleece was a better heat insulator than nylon, cotton knit, wool or denim. Again the workbook supplied a blank template modelled on the NCEA version. In a whole class discussion, Mr Ashcroft set the scene for the investigation by asking the students questions about their personal recollections of material types and their properties. He then turned the discussion to variables, drawing parallels between variables in the scenario and those in an investigation into the germination of seeds. After this introductory discussion lasting about five minutes, Mr Ashcroft allowed the students to continue with the exercise on their own. He attended to students who were off task, and later provided supplementary notes on the whiteboard, including background science concepts such as insulation and a prompt question: "What factors other than the insulating factor of the fabric could affect how warm/cold something would stay if wrapped in that fabric?" He provided a number of hints:

- "What about the thickness/amount of fabric used?
- My neck felt cold when I stood in front of the fan
- Does something at 100 degrees Celsius cool as quickly as something at 30 degrees Celsius if they are placed in a room at 25 degrees Celsius?"

For the remainder of the session he moved around the room, monitoring progress and assisting students when needed.

I took this opportunity to seek elaboration from my student participants about various aspects to do with their classroom experiences. I was interested to know more from Alex about the apparent 'learning relationship' that was occurring with his neighbour in class. I spoke to Alex and his neighbour together, and asked if they worked together all the time. Alex replied "Yes, we keep in step when we're completing tasks. If I don't know something I ask him and he tells me, and vice versa. We teach each other, because we are usually ahead of the others". He found the student workbook "okay, but it didn't have the model answers at the back". He liked to use model answers to check his work to see if he was "on the right track". When I looked through his workbook I noticed some of the set work was incomplete. When I commented him about the gaps he said “I only do exercises or homework that I think I don’t know - that I need to. I don’t waste too much time doing things that I already know how to do”. Alan and his partner worked consistently on the planning exercise for the duration of the lesson.

In a conversation with Carol and Anne, they told me how they really liked the NCEA type questions in their workbook and textbook for the practice. Like Alex they both agreed that models answers in the workbook would be very helpful with their learning “so you can self-check, especially at home”. Anne noted “when you go into exams it’s very good to meet exam questions that you have already experienced in a similar form”. Carol had quite a few gaps in her workbook and explained “I don’t have time because I have to go to my tutor for language lessons after school”. Anne’s workbook was still missing work she had missed while absent at the school production. I often observed her talking socially to her peers during this lesson, and when I asked her what she thought about the planning task they were asked to do in class she found it “very boring. It’s kind of easy”. Carol was not making much progress with the task either, and although she worked steadily on it throughout the lesson she seemed similarly indifferent.

Steve’s workbook showed that he had attempted most set exercises, and he found the workbook quite helpful for this topic. He made good progress with the planning exercise, but I did observe him engaging socially on several occasions for quite lengthy periods, usually with Anne’s cluster of students. He was unsure about how to tackle the data processing and interpreting section, and appeared not to have started on this part by the end of the period.

Mrs Smith, the second of the relief teachers, took the fifth lesson I observed two days later. I was unable to be present at the lesson the day before, but I was able to find out from the students that they had worked on a practical investigation from their workbook to do with rates of reaction. My absence meant I could not determine how much of this exercise was completed by my students, nor the nature of their involvement. The investigation was into the conditions that affect the rate of chemical reactions, in the context of the reaction between hydrochloric acid and marble (calcium carbonate). In this investigation the students were to do some preliminary exploratory work, observing the reaction and finding out the word equation for the reaction. They were also given a hint how they might measure the rate of reaction (by the addition of a drop of detergent). Then they were to plan, carry out, record data, process and interpret and write conclusion and discussion sections. There also was a set of questions related to the science ideas behind the investigation.

The content of the lesson Mrs Smith supervised comprised problems and exercises set from the first chapter of the students' textbook (Hannay et al., 2002). There was no formal instruction in the lesson. A significant proportion of the class took time to settle into the tasks, and the relief teacher frequently checked student behaviour by calling across the room. Carol, who had started her work promptly, was interrupted several times by the teacher's calls. The relief teacher spent the lesson walking around the room checking progress and behaviour, and helping students with difficulties. Alex spent the first 5-10 minutes in discussion with his usual partner. Some of the discussion was about the text, but Alex was not as committed to his work in this lesson as he had been in previous lessons. Off task behaviour of other students often distracted him, although his partner stayed busy for the bulk of the lesson.

Partway through the lesson I spoke to Alex about his progress, and he commented: "I'm not rushing ahead because my partner is a question behind". I noticed his notebook was very ordered and tidy, with headings in red and underlined. He explained "It's just me. It helps me revise. When I go through my notes these words stand out and it helps me learn".

Steve's neighbour had arrived late, without his textbook, so he shared text with Steve. Despite the relief teacher's attempts on several occasions to explain and support Steve in achieving the next step in the problem solving, Steve could not remain focused for long once she departed. He often consulted his neighbours, and barely managed to complete half of the set work in the period. When I asked how he finished such work he told me he sometimes "did it for homework". His friend added the comment: "It depends on the surf" and they both laughed. Steve's notes were well organised, but many questions in his workbook were unfinished and answers typically brief.

Very few students in the class worked consistently during this lesson, the most common activity being social interchanges. However, the two girls Anne and Carol were conspicuous by their prolonged attention to the written task. I noted a male classmate who Anne normally conversed with, was not in class, and she did not chat to any neighbours until the last quarter of the lesson. In fact, she earned the relief teacher's open praise for completing the most work in the period – she actually finished all the set problems from the textbook. At one point my attention was drawn to Carol's work by Anne's comment "You're writing the answers in your textbook!" Such 'abuse' of a textbook would normally be unacceptable behaviour for a student, but it transpired that Carol actually owned the textbook. She had bought the book the previous year and chose to write her answers in it. She saved her notebook for notes only. She wrote in her diary during the lesson too, telling me "I write things in there about words or the work that I cannot understand so I can ask my language tutor about it". Carol did not seek assistance from the relief teacher at all during the lesson, and she told me she normally only asked her teacher Kathy for help on a few occasions. She relied mostly on her tutor for assistance, and sometimes on Anne in class.

4.3.4.5 A practice climb on a neighbouring peak

Three weeks later, during which time content for another standard was taught, students were given the opportunity to test their skills in a situation very like that they were going to meet for the 'final ascent'. I could not attend the first planning session of the formative assessment on the Thursday, which students did individually, but I was present for the second lesson on the Friday when they carried out the practical work. The students had been asked to imagine themselves as research chemists employed by a company to investigate the rate at which their magnesium metal reacts with hydrochloric acid (see Appendix N). They were to plan and carry out an investigation, and prepare a written report of their procedure and findings, including a final recommendation on the best temperature for the reaction. This information was to be used in the company's new manufacturing plant.

Kathy had taken their plans in, marked them and on the basis of this assessment placed the students into mixed ability groups, but in different combinations compared to the formative assessment groups. Again my students were dispersed through the various groups. She instructed each group to talk amongst themselves, share plans and then write a shared or common method that they all agreed on. If there is no change in a particular plan the students could write on their script "see original plan". Kathy gave some last minute advice: "Be careful you have only 10cm of magnesium and a set volume of hydrochloric acid. Think carefully about your quantities. It needs to be written in your method". She ran through the report template, reiterating key points: "Every person has to write their results in … write the group method. Everyone has to have a copy of it … do the experiment as a group … record data in a table". The students had a third class period in which to finish the report. They could attempt to write the report up in this second practical session, but Kathy insisted: "Don't rush it". I decided to observe just one group (Carol's group), again for logistical reasons.

The student instructions contained advice about doing trial experiments in the planning stage to ensure the plan or method would produce the desired outcomes. Individual students would have done this trialling. I am not aware from their plan, reports or interviews about the extent of this trialling but it seemed, from Kathy's comments in our telephone conversation after the formative assessment, that some trialling did occur. She remarked: "A student wants 40mL of acid for practice. Imagine if every student wants this!" Certainly no trialling was done in the second practical session. I noted too that the student planning template did not require the purpose of the investigation to be stated, presumably because it was stated in the text of the scenario.

Once the common plan was recorded the group could start data gathering. Carol's group was made up of herself and two boys – one of the boys, Charles (a pseudonym), was Asian, and the other, Andrew (a pseudonym), was a New Zealand European. I learned that Anne was also to be a member of the group, but she was absent. The group began by quietly reading their returned plans, and then Charles approached Kathy to clarify the task. Kathy came over to the group and repeated that they needed to talk and write down a common method. They each begin writing, but there is little talking beforehand. Carol did ask a question of the other two, and Andrew provided an explanation. I noticed of the seven groups in the room, four began their experimental work immediately. I asked Alex why his group was going straight to the practical, and he explained: “They decided mine is the best … it’s the most detailed and they’re running with it”. Kathy returned to Carol’s group and enquired: “Are you changing the method? She suggested they refer to Andrew’s plan because “I’m worried you’ll run out of practical time”.

Andrew was the most confident of the three students in Carol’s group, and he led the way as they began to carry out the investigation. Charles was unsure and sought information from Kathy again. Meanwhile, Carol busied herself measuring out the magnesium ribbon and the hydrochloric acid, guided by Andrew who had drawn up a result table. Charles watched, communicating with Carol every now and then in a Korean dialect. I asked Carol if she had read and understood Andrew’s plan, and she replied “Yes”. I moved to Steve’s group for a few minutes and found they too had chosen to follow a single unchanged plan. It was the most detailed of the group’s various plans, and had not been read through by the group before they made their choice. Instead they were reading the plan step by step for the first time as they carried it out. Observing from afar I noticed that Steve was often not directly involved in the execution of the plan, and was frequently inattentive. Despite his small contribution to the group effort, he was able to finish the lesson with a full set of recorded results.

I watched Carol's group carry out their plan with few hitches in the time provided. Kathy interrupted the class at one stage to remind them that in the summative assessment they needed to be aware of timing because "you have two hours, but you must do everything". I asked Carol's group if they had learned anything so far from the formative experience, and Andrew commented: "To hurry things up". It appeared he was gaining procedural knowledge about assessment requirements. While Carol collaborated with Andrew to collect the data, Andrew was the group member who recorded the data and performed calculations. By the end of the session Carol, who tidied the equipment away, had not managed to record the results. Like Steve, Alan had managed to secure a set of recorded results from their practical work. All students handed in their scripts to Kathy at the end of the lesson, and received them back in the next lesson for the final write-up.

4.3.4.6 Debriefing after the practice climb

Students completed the report for the formative assessment on the Monday, and Kathy assessed the reports that night ready for return and feedback to the students on the Tuesday. Although I could not be present for the finish of the formative assessment and the feedback session, I did manage to secure several sources of data that gave useful indicators of events during these lessons:

- an audio tape of the feedback lesson, recorded by Kathy

- notes from a follow up telephone conversation between myself and Kathy after the feedback lesson.

I also had the one page "student journal" questionnaires (see Appendix O) for students to complete, as a record of their experiences and thoughts about the formative assessment. Two students, Alex and Steve, handed in responses to me but they did not contain any information pertinent to these two lessons. The audiotape of the feedback lesson revealed that Kathy had assessed the students' reports and provided written comments on each script. She used this lesson primarily to prepare the students for the summative planning exercise. Their summative planning, done individually, was to be held the following day, followed a week later by a two hour exam slot for the practical and report writing portions. Kathy explained:

> You'll be given a sheet like this (the formative assessment sheet), with all the instructions. It's the same, except it's not going to be the reaction between hydrochloric acid and magnesium metal. It will be something else. You'll have to decide what is the variable you'll be changing. Yesterday it was the temperature, so write it in your plan. Tomorrow it's not going to be the temperature you change. Remember what it could be? … That's right, surface area or potency – the scientific word is concentration. Very important that the formative was temperature, you're not going to get that for the real one obviously.

In the feedback session Kathy led students step by step through the report template for the formative assessment. For each section of the template she acknowledged those aspects they had successfully achieved in their reports, and those aspects that were missing or inadequately done. Where gaps and weaknesses were identified Kathy ensured students were informed of how to remedy the problem, either by telling them directly, or by guiding them through prompt questions to the next learning step: “How would you change the concentration of an acid … dilute it. Good! … When you add water what are you doing? … diluting it, making the concentration weaker”. She encouraged them to write down appropriate responses on their formative scripts if points were missing. She also gave out information related to assessment procedural knowledge, and exemplars. For example, when discussing the dependent variable she asks: “How will you measure that?” A student responds: “Stopwatch”, to which she replies: “Now you can’t just put stopwatch, see how much room you’ve got (on the template) … put ‘Start the stopwatch when you put the magnesium into the acid and stop the stopwatch when it completely dissolves or disappears’. Write that in now”. She surmises: “Tomorrow you’re going to be timing how long something takes to dissolve I imagine. How will you do that? The same way as for this experiment?” Sometimes she gave quite targeted tips: “I suggest in the processing you find the average and draw a graph – that’s my suggestion”, while in another instance she is almost directive: “Make sure you’ve got your results table memorised. The table was given to you in the formative. You’ve got no direction for the summative but I suggest strongly that you do a table like you’ve done today”.

Kathy continued in a similar vein for the whole formative assessment template, giving instructions on how to present full and accurate responses. Most advice and guidance were directed towards the early sections related to the independent and dependent variables, the measurement of the dependent variable, the control of other variables and the method: “Lets talk about your method. You need to explain step by step what your plan is and to get an ‘excellence’ your plan has to be workable. Someone else should be able to carry it out without a hassle. It needs to make sense”. She gave them credit for writing good methods on the whole, but pointed out that the following day they were going to write individual plans: “Then you’re going to go I imagine and talk about it, and when you come in next Friday to do the group one, don’t just pick the one that looks the longest. Make sure you read what they’ve written, pick out the best from each method, combine them and learn from each other’s method. Share each other’s ideas and write the group one”. Each student was to provide a copy of that group method in their report, although if an individual group member’s original plan was to be used as the group plan the degree of detail required in each group members report was unclear. Kathy’s instructions on this facet of the report were to put in numbered points because: “You still need to give a copy of the method, but for further details see the original one”.

The audiotape recorded many instances during the teacher-led discussion, of high noise levels amongst the students. Often Kathy’s voice could be heard trying to speak over students’ conversations and she had to call for quiet or on task behaviour frequently: “Come on guys, you’re not on task. I’m very disappointed”. The noise levels became quite intrusive and were probably indicative of a degree of disengagement from the question-answer activity. Kathy commented: “You should be paying attention. You have gone through three lessons of planning, carrying out the plan and writing it up, and for some of you the work in front of you is not very good at all. Tomorrow you’re writing the real plan worth four credits so none of you have the excuse to talk”. In her telephone conversation with me Kathy expressed her frustration about this lesson, and the growing amount of time and resources she felt was being spent on this achievement standard: “I don’t know what learning we are getting out of it! I felt good teaching it, and I did mark it [the formative assessment]. I wrote all over it and I spent a whole period going over it. They just wanted to ask me everything; they didn’t want to think themselves. I felt annoyed about it. It’s costing the school a lot of money – some students are wanting large quantities of chemicals for practice”. Her disappointment stemmed largely from the attitude of students towards achievement: “I thought they would be focused and keen to learn what to do for the summative. It’s strange, they just want the credits – they have this idea that that’s all they need”. A month later, in her final interview with me, her frustration with the feedback session was still evident, but she was beginning to rationalise the students’ behaviour commenting: “Maybe it was because they felt they had done it well and they knew it and they were going to be fine”. She admitted her disappointment stemmed in part from her own desire for the students to get the very best achievement in the summative assessment: “I felt that it was so valuable and I didn’t want them to do anything wrong in the summative.”

One incident from the tape did suggest that substantive learning was occurring for some students. Kathy had asked students to copy an exemplar statement for the conclusion section of the formative assessment. It read: “By increasing the temperature the rate is decreased”. She had been making the point that since the term ‘rate’ was mentioned in the purpose of the investigation it must be alluded to in the conclusion statement. She explained rate as “the time or how fast it takes for the magnesium to dissolve”. A student queried Kathy about this statement: “Isn’t rate the speed of something?” Kathy maintained her stance through the ensuing discussion with the student that the time taken is equivalent to the rate, but when another student asks: “Is rate the speed at which it happens?” Kathy reconsiders. She replies: “Yes, you’re right. Just as well I’ve got some people onto it”, and she corrects the conclusion statement for the class.

I was able to retrieve three formative assessment scripts from Alex, Carol and Steve, including the Student Instructions Sheet. Their assessed papers showed feedback from Kathy in the form of ticks for appropriate responses, and written comments including corrections and prompts like “how?” or “recommendations?” There appeared to be no indication on these papers of their level of achievement. Alex’s report showed no corrections, but Kathy had provided four suggestions for improvement, one of which was grammatical. Of the other three suggestions, two related to giving more detail when describing how variables were to be controlled or measured. The other suggestion reminded Alex that a recommendation to the mock company on a suitable temperature for the manufacturing plant was required in the task specifications. Looking more closely I noted his report showed Alex had written comprehensive, in depth and accurate coverage of most aspects of the investigation. I did observe a few inaccuracies related to procedural knowledge in the data gathering and processing areas – his table of data showed units within the table and a mix of units in given columns or rows, and he used a histogram where a line graph would have been more appropriate for identifying trends and patterns between the continuous variables of temperature and time. Kathy had not commented on either of these discrepancies. However, his conclusion and evaluation section demonstrated he had a deep level of understanding of the concepts and procedures involved. He clearly linked the experimental findings to the background science concepts, and made appropriate suggestions for refinements to his method: “The reason why magnesium ribbon reacts faster in warmer acid is because the acid has more energy, so the particles move quicker and collide with the magnesium more frequently and so speeds the reaction up … I would do three trials of each temperature at the same time to save time". Alex's grade, in my opinion, would have been very close to excellence considering the generic requirements of the standard.

In his journal entry Alex was positive about the formative assessment experience: "It was useful in preparing me for the real thing. It helped me study and know what to expect from the real one. I have learnt how to structure experiments and fill in the sheet with the questions". His instincts were that he was achieving well in most areas, but he thought he could improve his achievement by using examples or actual results in the conclusion part of his write-up: "My conclusion usually says that one is 'better' or 'quicker' than something else, rather than actually stating the figure, for example, 20 degrees Celsius is hotter than 15 degrees Celsius". He was confident about the upcoming summative assessment because "I find it relatively easy to understand 'science' concepts, and for learning I just read through all my notes and do exercises out of the book. I learn very quickly by reading things". In his final interview Alan again expressed the opinion that the formative assessment prepared him well for the summative: "Yes it definitely prepared us because she said it (the summative) was very similar, so just remember what you are doing".

Carol's formative script showed more interventions from the teacher, including correction of the all-important dependent variable. Under the requirements of an exemplar NZQA assessment schedule for the standard, correction of this error would have given Carol a non-achievement grade for this section. Where Carol was to specify which variable was to be measured in order to get some data from the investigation she had incorrectly identified the size of magnesium as the dependent variable. However, in answering the following question about how this dependent variable was to be measured she wrote “will be measured by stop watch to see how fast it dissolves”, indicating she understood time was the variable to be measured. This understanding also came through clearly in her method. In such a case assessors would award Carol achievement with merit, because on balance she had demonstrated correct application of the dependent variable concept and her plan was ‘feasible’. Her plan was unlikely to be judged ‘workable’ (for excellence) because detail about how other variables were to be controlled was not complete, such as the amount of magnesium to be used, and it was unclear how the data was to be processed to obtain information about the rate of the reaction.

Carol’s difficulties with English affected the clarity of some of her statements, but she was able to communicate most key points successfully. She used diagrams throughout her method, almost as a form of hieroglyphics, to illustrate and depict items of equipment, quantities of materials and techniques. This strategy was very effective for improving her ability to communicate ideas and her method only required minor adjustments to reach excellence standard. Sometimes Kathy indicated where information could be more appropriately placed by the use of arrows. Carol had copied down some of Kathy’s suggestions onto her formative script from the feedback lesson, including descriptive phrases and a further diagram for the method. She indicated in the second interview that Kathy’s written comments often served as useful reminders and helped her feel more confident about the final assessment.

In the report section of Carol's formative assessment there was a reference that indicated the group had adopted the method of one student as their group method. The results table showed repetition of experiments and averaging of results had occurred. Like Alex, Carol had placed units within the table and had drawn a histogram for the processing part, showing the results of all trials rather than the averages. Unit labels were missing on the axes. Kathy had not made comments on these points. In the conclusion and evaluation sections Carol's responses lacked detail. She did not link the findings back to the purpose of the investigation, which would have automatically cost her achievement at merit level. While Carol did not attempt an evaluation, she did make appropriate links with the background science: "Hot molecules moves faster so when we put the magnesium it dissolves faster, but cold molecules doesn't move very much. So magnesium dissolves slower". On the basis of her report, Carol would in my opinion be likely to receive an overall grade of achievement for this formative assessment.

Steve's formative plan and report also showed a number of corrections and comments from Kathy. He was successful in identifying the independent and dependent variables, but only managed to describe how to manage the measuring of the dependent variable of time. He was unable to isolate other relevant variables. His method failed to provide anything further on the variables, and there was insufficient detail to be judged feasible. In his report Steve referred to the use of another student's method for the obtaining of data. He was able to record sufficient data in an appropriate table to meet reliability requirements and obtain averages. Like the reports of Alex and Carol, units appeared inside the table and Steve constructed a histogram for the data processing. Kathy commented on neither point. Steve's conclusion was based on an interpretation of his results: “That the more the acid is hot then the more faster the magnesium boils/bubbles”, and probably contained sufficient information to warrant achievement. Steve had not corrected Kathy’s comment about rate after the debate in class, and her written restatement of his conclusion statement still read: “By increasing the temperature of the acid the rate of the reaction between magnesium and hydrochloric acid is decreased (time has decreased)” in his script. He did not attempt to link his findings to the background science, and his evaluation, which read: “I would make the temperature better by maintaining it”, needed further explanation for the reader to understand the reasons for his suggested change. In the final interview Steve appreciated Kathy’s comments on his formative script because they reminded him not to repeat similar mistakes in the future: “It’s a kind of reminder of not to go too far out of what you’re actually writing, because sometimes I might go a bit further than what I have to write”. He felt more confident about the summative assessment as a result. Steve’s formative exercise would have earned an achievement in my judgement.

Anne did not provide her formative assessment script for my data collection, and since she was absent for part of this exercise I was unclear whether in fact she had submitted a report to Kathy. She did comment though, in our last interview together, that she found the formative assessment useful for her learning: “Once we had done the practice we all knew what we were doing”.

4.3.4.7 Ascent under cloud cover

As mentioned earlier it was school policy not to allow observers into the summative assessment session, so I was unable to witness the ‘scaling of the peak’. Despite being unable to observe the students perform the investigation, I was able to obtain copies of the students’ summative assessment scripts, but not the assessment schedule. Grades written on assessments showed that they had all achieved the standard, with Alex and Anne achieving excellence and Carol and Steve achieving merit. These results were evidence of further learning for Carol and Steve because each had improved their achievement by one level, and Alex had maintained and confirmed his achievement level at excellence. Anne also achieved at excellence level, but without a formative assessment grade to refer back to I cannot credit her with improved learning on the basis of this indicator alone.

The summative assessment task followed the same pattern as the formative assessment. The scenario this time centred on an antacid tablet used to relieve indigestion pain due to excess stomach acid (see Appendix P). The ‘research chemists’ were to investigate whether the surface area of the tablets affected the rate at which the indigestion tablets Painaway dissolved in hydrochloric acid, and present a report of their findings.

While Alex’s formative assessment met the criteria for excellence, his summative report was far more comprehensive and detailed than his formative attempt. The report contained full, sequential accounts of experimental methods with labelled diagrams tabulated and processed data; and coherent, relevant interpretations and explanations. Reliability issues had been addressed appropriately. Along with these successfully achieved aspects, it would appear that his effective linking of the experimental findings to the purpose of the investigation in the conclusion section and recommendation to the company, and his evaluative comment about a possible improvement to his experimental method earned him achievement with excellence. These statement received 'ticks' from the assessor, although his evaluative comment was restricted to one sentence: "If I were to do my experiment again I would make sure that I dried the beakers out after I washed them each time, which would make my experiment more accurate". He did not explain why this action would make the experiment more accurate. However, his explanation of the findings in terms of relevant background science concepts was thorough and well constructed: "The theory behind why surface area effects the rate of reaction is, that the greater the surface area the more particles there are which are exposed to the hydrochloric acid, which means that the hydrochloric acid particles are more likely to collide with the Painaway tablet particles and collide with them more frequently which increases the rate of the reaction". He then accurately explains the converse, when there is less surface area, to enhance his explanation.

I noted several procedural gaps and errors in Alex's report that appeared not to be taken into account when considering achievement for excellence. When identifying a variable that needed controlling to make results more accurate, Alex appropriately identified the temperature of the acid. However, where the report template required Alex to explain how this variable was to be controlled or measured he wrote: "The temperature will be measured using a thermometer, and it will be at room temperature". He did not explain how to keep the temperature at room temperature should it deviate. A procedural error from the formative assessment recurred in the data processing section of Alex's summative report. He repeated the use of a mix of units for his table of results, for example, 1min 12s rather than 72s or 1.20 min, and they were placed inside the table rather than with the quantity headings at the top of columns or side of rows. Realising that Kathy had not picked up on these points in her feedback to students, I referred to the textbook and workbooks that students were using in class to check for this procedural practice. Both reference books clearly showed units in the headings for table columns and rows, and only one unit per column or row. This was a piece of procedural knowledge that Alex was not acquiring, either from teaching or his own studies and research. Again he drew a histogram for processing his results, although on this occasion it was more appropriate than a line graph since the surface area variable would be difficult to quantify. This horizontal axis was labelled “Type of tablet”, with each column bearing the labels “Whole tablet”, “Halved tablet” and “Crushed tablet”. The graph identified a relationship between the surface area of the tablet and the time taken to complete the reaction, and hence the rate. Alex's practice of using histograms is interesting because his references materials both advocate the use of line graphs “trends and relationships between quantities are displayed on *line graphs*” (original emphasis, Hannay et al., 2002, p. 11), and Kathy does not appear to have taught differently.

Anne's paper achieved excellence also, but possible differences in the nature of the learning when compared with Alex's report were indicated. Differences were evident in the suitability of some aspects of experimental design; the extent of descriptive detail and accuracy of language use; and the depth and range of interpretative and evaluative comments. The first obvious difference between the thinking of Anne and

Alex lies in the choice of the variables to be controlled. Alex had chosen concentration and volume of acid, the temperature of the acid and the type of container, while Anne had chosen volume of acid, weight of tablet and size of beaker. Alex appears to have accounted for the factors that have been repeatedly recognised in the theory and formative work leading up to the summative assessment as affecting the reaction rate, namely temperature and concentration. Anne has not acknowledged either of these two factors specifically as needing control, although in her method she talks about testing the temperature of the acid using a thermometer and recording the temperature in the table of results. In her initial plan when describing the temperature aspect she did include the phrase "make sure it's the same for all experiments to follow" but not in the final shared plan in her report. A column was provided for temperature in the final results table but no data was recorded. By choosing volume of acid and mass of tablet she could be demonstrating some understanding of the importance of concentration on rate.

In terms of procedural knowledge, it would also appear that Anne has not taken into account certain ideas to do with assumptions. For example, it could be reasonably assumed that since Painaway tablets are commercially produced on a large scale to precise specifications of mass content per tablet, each tablet had the same mass and that control of mass was therefore achieved. Furthermore while Anne performed three runs for each surface area variation to improve the reliability of her results, her requirement to "control" the mass of the tablet indicated a lack of awareness on her part of the role repetition plays in addressing the systematic error that using individual Painaway tablets might represent. However, I have no evidence to suggest that Alex had this understanding of systematic error either, which is not surprising given that errors did not feature highly in Kathy's intended curriculum. Like Alex with his instruction for controlling the temperature variable, Anne's instruction of "weigh them on the electronic scale" for controlling the mass of tablets would allow the investigator to check if the size of the variable was remaining constant, but did not inform the reader how to control the variable should its value change.

In other instances Anne used incorrect (even invented) vocabulary, such as "time of dillusion" to mean "time of dissolving" when describing the dependent variable. She used the term "dialuted" when describing how timing was to be carried out, apparently to mean "disappeared" because she supplied it in brackets after the term: "Stop the stopwatch when the tablet is fully dialuted [disappeared]". Later in what appears to be an addendum to the method, or a rewrite on a loose sheet of paper, she writes: "When the tablet has fully dissolved stop timing straight away". In this instance she has used the appropriate vocabulary, not "dialuted".

In her data collecting and processing sections some anomalies appear. Data for one trial (the powdered tablet) is recorded to two significant places after the decimal point, but not for the other trials. The only unit present is 's', which presumably refers to seconds. This unit is found inside the table and is only present for two of the nine measurements. Averages for the three trials of whole tablets, chunks of tablets and powdered tablets are "57.33; 40; and 1.18s" respectively. It is unclear to the reader if the last entry is 1.18 seconds or 1 minute 18 seconds. However, when you examine Anne's data processing calculations and the histogram graph it is clear that she has read it as 1 minute 18 seconds. The data as it stands in her calculations and graphs shows an anomalous result for the powdered tablet, since the expected result would be that the powdered table gave the shortest dissolving time and not the longest. Anne acknowledged this anomaly in her evaluation and attempted to explain it with some allusion to the background science: "We believe that the powdered should have got the best time because the powder should have dissolved fastest in the acid. The chunks could have possibly been the fastest because it was easy to breakdown and dissolve". The links are not adequately explained, and if the argument was to be accepted in this form there is even more reason to expect the powdered tablet to dissolve fastest. At no stage did she suggest that there might have been an error of some kind. An explanation could be that Anne's group mistimed, or misread their data, and that the reaction time should have been read as 1.18 seconds. That would mean a very fast reaction rate, but more in keeping with expected results for the powdered tablet. I crosschecked with Carol's script, since she worked in Anne's group for the summative assessment, and found the same numerical quantities in the results table but confusion over the units. Their histograms showed similar patterns. Another minor discrepancy in Anne's data processing was the lack of a heading on the histogram graph.

It is in the conclusion, discussion and evaluation areas that the greatest difference in the nature of the achievement can be discerned between the reports of Anne and Alex. Kathy had awarded two 'ticks' in this section of Anne's report. One tick was linked to the stated recommendation, which accurately reflected the experimental findings that: "Chunks proved to be the best tablet to use, so my recommendation is chunks. We originally thought that the powdered tablet would be the best to use but were wrong". The other tick related to a statement about limitations: "One of our limitations was that we didn't have electronic scales or a thermometer to use". Anne did not comment further on this point, so it was difficult as a reader to interpret why this was deemed a limitation. The student instructions for this part of the report (conclusion, discussion and evaluation) tell students that they "will include interpretation of the data and should include mention of any assumptions you have made and any limitations you have identified. It will also include a final recommendation of the company including any discussion on the results and conclusion that may be necessary. You should also link your results to the science behind the investigation". It seems that Anne's inclusion of a recommendation and mention of limitations had been suffice to earn her achievement with excellence. While Anne's report met the excellence criteria as interpreted under the science department's moderation process, there are obvious indicators that Alex's report represents deeper levels of understanding and skill. It could be argued that in light of the excellence criteria in Science Achievement Standard 1.1 (see Appendix A), where a student is to "provide a comprehensive evaluation or discussion of the investigation", that Anne's report did not reach excellence standard, particularly in the interpretation and evaluation section because her explanations were not convincing on theoretical grounds, nor were the reasons for her evaluative comments made clear.

As mentioned earlier Carol worked with Anne for the summative assessment, and consequently their data were identical. Carol achieved with merit although her method and data collecting and processing sections were sufficiently detailed and accurate enough to be achieving at excellence level. Her conclusion, discussion and evaluation, however, received no achievement acknowledgements by way of 'ticks'. She did make a statement that showed a link with the purpose of the investigation: "Chunked Painaway tablet is dissolved faster than whole or grounded tablets" but her attempts to explain the unexpected result that chunked tablets dissolved faster than powdered were unsuccessful in terms of the background science. Her initial comment "because whole tablet has got no space between molecules and grounded molecule has got too many spaces" contained contradictory ideas, and she continued these contradictions throughout the rest of the discussion section using a combination of diagrams and descriptive phrases. She successfully accounted for the whole and chunked tablet results by illustrating the degree of access that acid particles had with tablet molecules, but could not account for the powdered result using the same rationale. Like Anne, she made no suggestion that there may have been an experimental error. Her English limitations sometimes made phrases difficult to comprehend, such as her description of the behaviour of the grounded tablet: "Too many powdered tablet. It sticked each other in the acid so got heavier and slowest dissolving". Her discussion was hampered by her lack of procedural knowledge about the possibility of experimental error, and her inability to express ideas concisely in English.

In one respect Carol's write-up of her method was superior to that of Alex and Anne. Unlike Anne, she had retained the idea of temperature being a variable to be controlled from her initial plan, and specifically identified it in the planning template. She then explained how temperature could be kept constant by "putting ice or hot water around the beaker", which is an aspect of experimental design that both Anne and Alex omitted. Despite Carol's inability to gain above achievement level in the final section of her report, she gained an overall grade of merit for the standard. This is likely to be another 'on balance' assessment decision because of the high calibre of Carol's responses in the first two parts of the report.

Steve's paper was interesting not only for what it revealed about his capabilities and scientific understanding, but also for what it revealed about the evidence required for particular assessment decisions, and perhaps the conditions under which the summative assessment was conducted. Although his initial plan was a little wordy and repetitive in places, it contained most key elements of a workable plan, which would earn Steve achievement at excellence level. The only point missing was how to maintain the temperature of the reaction mixture at room temperature. However, when Steve came to fill out the report template during the summative session his descriptions and explanations about specific aspects of the experimental design lacked clarity and were quite confused in places. When identifying the independent variable he wrote: "Investigate the rate at which the indigestion tablets dissolve in dilute hydrochloric acid. And surface area of Painaway tablets". Later when identifying other variables that need to be controlled he included surface area again, and when describing how to control the variable he wrote "You should change each tablets surface area so it is a fair test and accurate, for example, crushed, powdered and normal". Another variable he mentions for control is "measurement" commenting: "You should decide on an accurate measurement for each tablet for example one cubic centimetre". It seemed through these confused statements and errors that Steve was negating the precision and clarity of his original plan. Since I was not able to observe the group in this situation, nor was any other data available I can only speculate that perhaps group opinion resulted in changes to Steve's thinking, or the pressure of the summative situation caused confusion on Steve's part.

The group method that Steve recorded in his report also showed discrepancies compared to his original plan, resulting in a less coherent set of instructions. Instead of advising the reader to repeat experimental steps for each of the three different trials, step four of the eight step set of instructions read: "Crush, powderise or leave normal the tablets with the mortar and pestle or with a blade", and step eight read: "Repeat test three times so it is accurate and make it a fair test". From these statements the reader could infer that the experiment involved three runs in total, compared with nine runs in Steve's plan. The group's method lacked some of the clarity of Steve's plan, but their result table had nine spaces for recorded results. Presumably the group intended step four to represent three trials, not one.

In the data gathering and processing section units appeared within the table, and the histogram had no title. The graph's horizontal axis had no label and but each column was labelled as trial 1, 2 or 3. The final interpretive section contains just two statements. One statement is under the conclusion heading: "That the powder is the quickest to dissolve and I recommend this form", and under the discussion/evaluation heading: "This form as a powder is good because it is fast compared to the whole and 2 halfs. And it would be easier to swallow". Steve has provided a recommendation, but there are no links with the background science, nor an evaluation of the experimental method. It would seem because Steve has mentioned the independent and dependent variables in his conclusion statement the link has been made with the purpose of the investigation, and hence part of the criteria for merit had been met. There is no interpretation based on the data obtained though, such as "as the surface area of the tablet increases, the time taken for the reaction decreases", which is the evidence required for achievement by the standard. It would seem that Steve had not achieved the standard in this segment. The assumption is that the assessor (Kathy) has made an holistic decision to award the standard at merit level based on Steve's performance in the first two areas of the report.

4.3.4.8 The mountain's guide's reflections on the climbers' achievements

As Kathy reflected back on the progress of the students over the three to four month time span, from their first introduction to the investigation standard to their final achievements in the summative assessment, her initial impressions focused on their gains in procedural knowledge to do with assessment:

> Well, they learned in that we gave them a formative assessment … if they didn't do something that was required in that formative assessment, we could point out to them and show them this is actually what they have to say or what they would have to write, or how they would have to carry it out … by the time it got to the summative, the majority of them were picked up on those things and actually realised they had to do that and did it correctly in the real one.

She also focused on the concepts related to the science contexts in which the investigations took place, and saw the investigations as a vehicle for consolidating these concepts: "Well, we looked at rates of reaction so they have picked up the four important factors that affect rates of reaction … so I think those concepts, the science behind all that is pretty well gelled". As far as specific skills such as measuring or using a stopwatch "I haven't taught them anything new in that area". When I probed about students learning about investigating she felt initially that learning gains were not overly significant, since students came into Year 11 having done science fair investigations in Years 9 and 10. She did acknowledge that her Year 11 students could not write good plans and methods at the beginning of the year but said “they do it now because it’s sort of told to them in a formative way and then a week later they do the summative so they remember for the fact that they are going to get credits”. She felt the learning was short-lived, for the purpose of gaining credits for NCEA. All students in the class, bar three, received achievement with excellence, but Kathy questioned the depth of the learning, speculating that if the students were required to repeat the performance later in the year: “They would still have difficulties planning an experiment … I mean the mark says that they can do it, whether they could do the same thing again, I don’t know”. Despite her doubts about the extent of the learning, she did consider that learning had occurred: “I’m doing the same thing now with my Year 10s for their science fair … and they do get better at it but it’s still not perfect”.

Kathy had been present during the summative assessment, and she reported that students generally worked more cooperatively when deciding on the final experimental design. She reiterated her practice of reading through individual student plans ahead of the practical session, and putting students into groups “with students who wrote similar plans to each other, or if someone wrote a plan that was off the wall so to speak then they would need to be with someone who had written a good plan so they could at least come up with a plan that would hopefully work”. For the group plan Kathy insisted that each student rewrite it on his or her own script. They could not choose and refer to someone else’s plan simply because “it’s the longest and it looks like it’s the best. They have to read it and agree on it”. Kathy witnessed students reading one another’s plans and collaborating as they formulated and rewrote the group plan: “A few of them were sharing so they could all see it like a group of three, they would put it on an angle so they could all see it and then copy a bit of it and then add a bit here and there".

On reflection, Kathy believed the most effective strategies she used for promoting student learning and achievement in the standard included:

- her ongoing questioning to probe students' levels of understanding; "because it gets them thinking and they have actually got to reply to you"
- group work
- going through the actual standards step by step saying "this is what's required of you" and "this is what you need to do in order to get your credits"
- revisiting prior learning, and using exercises to remind students of "the language they had to know like the aim, what are variables, what are controls, all those sort of things … that experience that they have had would help them"
- practical exercises practising specific skills "because they enjoy doing practical work and it actually gets them thinking hopefully about what's happening and what they are testing", and
- making specific rather than generic marking schedules available to students for practice planning and reporting exercises, because being able to "see the exact answers" gave students exemplars they could refer to: "You verbalise it but it's always good to have it in writing I think and take it away with them".

When I prompted Kathy about the effectiveness of peer assessment as a strategy she identified benefits for able students: "I used it last year with the extension science class and I think they get more out of it than mainstream classes because they actually talk to one another and act as teachers … they say the things I would say". She admitted in mainstream (mixed ability) classes "a lot of them prefer the teacher to mark it. Whether they think the teacher's right or whether it's just that they can't be bothered, I'm not sure".

Kathy felt the summative assessment went well for students. The departmental decision to shift the teaching and formative assessment timing closer to that of the summative assessment was a wise move in her view: "It needs to be done closely". Keeping the contexts of the formative and summative investigations in the same area of science worked to the students' advantage because there was "not that leap of science, the difference in ideas … they are not having to change their track of thought". In hindsight the timing of the summative assessment was not the ideal: "It was the last slot on the Friday so we worked from 2.20 to 4.20 … all they wanted to do was go home … they had friends who had totally finished exams". Despite this difficulty the students did the experiment, and as Kathy observed "everything that I thought they should be doing". The panel marking and moderation system using a common assessment schedule made the task easy to mark and Kathy commented: "You could tell very easily who knew what they [the students] were doing and who didn't". The teacher in charge of Year 11 Science checked marked several papers per class and students had the right to have the marking reviewed if they so wished. The comparatively poor internal exam results for externally assessed standards and anecdotal comments from parents and students led Kathy to believe that students put most of their energies into the internally assessed standards at the halfway point during the school year. In her view students didn't see the practice exams "as being a valuable exercise at the time because it wasn't worth any credits". Her perception was that "the students are definitely under more pressure when it comes to assessment throughout the year than they were with School Certificate, and it is hard for some of them".

I asked Kathy if she believed her students saw her dual role of teacher and assessor for qualifications as an issue or tension. She did not know, although when students asked her if she was marking their summative assessments she had a feeling "from their body language that most of them wanted me to mark their work. They would perhaps feel more confident thinking that perhaps I was marking it". She thought "the dynamics of the school is that the students know the staff and we know them and the student-teacher relationship is brilliant, so I don't think there is anyone that would have thought, oh no, I don't want that person, or if they had thought that person was marking them they would have to try harder or less harder".

4.2.4.9 The climbers' thoughts

I asked my participant students in their second (last) interview to give me their impressions of their overall performance and learning experiences. Their initial recollections paralleled Kathy's impressions by focusing on the procedural knowledge they had gained. Steve recalled: "We were taught how to write out the procedures and aim, all the way through to conclusion … and actually doing the whole procedure. Just learning how to sum up everything into a conclusion and actually writing up the steps so that someone could easily just follow them without even thinking". Steve recalled having learned about doing investigations earlier in his schooling, but he had been surprised that now he had been "taught how to do it properly, not like actually sum it up with only a couple of words in your conclusion". He had not expected to learn how to write up investigations in such detail and he was pleased with his final grade. When I asked Steve how did he know that he had learned how to do investigations he said: “You know because when you go and do something in the experiment and you haven’t actually written the experiment, you are thinking what am I doing, but once you have written it up you can just motor along”. It appeared the act of writing up the experimental methods had assisted his understanding. He knew he had learned something “if we’ve been taught and you find it easy to do”. Carol believed that “I have learnt how to write aim and purpose of the investigation and how to carry out experiment” and like Steve she had not expected to learn to this extent. She was not satisfied with her merit grade, as she had wanted excellence. She attributed her inability to achieve excellence to not understanding the question and needing to “write more appropriately and correctly”. She acknowledged her limited English was a factor but also that “I don’t have many science knowledge”. She thought she needed to study more.

Alex commented too that while he had learned a great deal about doing science investigations in his earlier schooling (in Australia), and that he thought he understood how to do them: “I knew basically everything before … it’s more like recapping”, he had learned to approach them differently: “I never knew how to set it up kind of thing, the tables, key variables … I’ve never done it like that before. We just had a method”. He felt perhaps aspects of the investigation like purpose and variables “might have been known as something else, they might not have been known as variables”. He knew he had learned this different approach simply because “I didn’t know it before”, and now he understood it and was confident in its use: “I think it’s relatively easy that’s why we all got excellence”.

Anne reported too that she knew many of the basic skills but: "I've kind of advanced on those skills, like with the variables. It's like thinking what variables are there? What are they? How can I use them in this experiment? … there's more thinking". She comprehended that her procedural knowledge about experimental design and reporting had grown: "I understand them better now … I now know how to write it up, how to write the method properly and how to write quite precisely and exactly what you want to say". Anne knew when she had learned because "after you have learnt it you are kind of confident, oh yeah, I know that and you get the concept right". She was not surprised by what they learned because the lead-up teaching sessions had involved "heaps of practice and "when it came to the actual thing (the summative assessment), it was like we have done this a million time before, it was pretty much what we expected".

There was some variation amongst the group in their opinions about strategies that helped them learn. Strategies ranged from instructional approaches that the teacher employed to techniques that the students themselves used in given situations. Carol was very firm in her view that the teacher helped her by "showing and writing notes". Steve also found notes, especially definitions, helpful for his learning because "you've got something to look at if you can't ask your neighbour and you're home and you don't actually have a teacher in front of you". Alex was generally non-committal about most of strategies identified by his peers as helpful for their learning, but he did find reading his notes useful: "That's what I usually do, read things through".

Steve had found it most helpful for his learning when Kathy put them into groups to do the experiments because “you could feed off each other … some of us were better than others at some things and the others would be better at something we weren’t”. Steve was not sure why they were put into different groups for the formative and summative tasks: “Maybe to check our social skills or something … just to get some variety”. In the summative assessment he commented: “I think we worked a lot faster than we’d do in class … we actually knew what we were doing”. Anne found the group collaboration very helpful in the final investigation too: “We wrote a group method, we combined all our three methods and chose one we thought would work best for us … it was way better, it felt so much more easier because we had combined everyone’s knowledge”. She added that: “It was pretty much my plan that they all used, but we added in an extra step that I had accidentally missed”. She didn’t mind that the rest of the group followed her ideas: “It just worked well like that”, and they got on very well together because “our teacher purposefully put us in groups where our methods were similar or the same”. Anne recognised that Carol, who was in her group, needed some assistance: “Carol knew what she was doing but she was confused with all the English and she found that a barrier”. Alex thought that working as a member of a group was an advantage, because it’s “easier to do the experiment with more people”, although when people have different points of view he noted conflict can arise and “conflict is bad”. He would prefer to choose the members of his group because there’s “less conflict”. In his view, there was no advantage to him when it came to improving the experimental design for the summative task because the group had used his method unchanged. Carol was not so keen on the group work she had experienced for the investigations, even when her group members were speaking Korean, because they “were just talking about nothing”, and not science. She thought it would be helpful working in groups, but only if the group members "have knowledge about science". Steve made an interesting remark about the degree of challenge in the summative task: "It wasn't such a hard thing to think over. It's just crushing and grinding … it's demonstrating what's in our stomach". He perceived the effect of hydrochloric acid on an antacid tablet an easier reaction to interpret than the effect of hydrochloric acid of different concentrations on magnesium metal.

Anne reiterated that the constant practice of doing investigations and using the blank templates was hugely helpful to her learning: "We knew what we were doing, we just went straight into it so just by going over something and getting it precise before … it was a big help". She identified the practice sessions where an assessment schedule was made available very helpful, but not so the peer assessment aspect: "Not your mate marking it because they don't really know what they are talking about, they probably know nearly as much as you, but when the teacher went over it, it was quite helpful". Anne did acknowledge that talking the schedule through with a peer was useful, but "they are not some scientist, they would not really know exactly, but our teacher she has done the degree and stuff so listen to her". Anne clearly perceived her teacher's knowledge and capabilities as superior to her classmates. Carol and Steve found the practice exercises very useful too, especially having access to the marking schedules because as Carol commented: "We know what we got wrong". Steve agreed, adding: "Yes, because it was like preparing us for the exam … they are always helpful, we get them in every subject … it gives you a fair idea of what you've got to do and you just refer back to it".

Alex and Anne expressed some dissatisfaction with the student workbook from which Kathy set work. Anne found some of the early exercises boring because “they’re obvious”. She had covered the concepts and skills in previous years and she considered exercises like writing definitions not very helpful because they are “common knowledge”. She confessed that some items she “never actually bothered to fill out completely because they are a bit boring”. Similarly, Alex didn’t find much challenge in many exercises, especially data processing, which he found “pretty easy, you kind of learn that in maths”. Anne occasionally found the work confusing: “Sometimes the questions are completely irrelevant … they don’t relate specifically to what you are learning and so we learn something different to what the questions tell us and sometimes they are really confusing”.

Several students saw study as a means of improving their learning. Carol had already identified that she needed to do more study to improve her grades, and Steve recognised that “study is a good one, especially when it comes to exams, and I think maybe going over what you’ve done at the end of the day would be a good idea”. Steve admitted that he didn’t revise work but could see “understanding is the main thing, otherwise you go into an exam and you don’t know what you are doing”. His main method of learning was “just listening and thinking”, and when he didn’t understand something in class “I usually get a mate who does know what they’re doing and I just ask him”. When his peer did not have the answers Steve was prepared to ask the teacher but this didn’t happen often. Carol relied heavily on her out of class teacher, but she did ask the teacher sometimes and found her approachable.

The group still had mixed feelings about NCEA. Steve was comfortable with the system, adding: "It's good not having too many assessments, it's enough". Carol on the other hand felt nervous: "I've got everything in my brain", but when it comes to answering the questions she explained her brain "just blanks". Alex was not in favour of receiving credits: "It's kind of stressful, makes it sound like university". He preferred the traditional system of exam percentages: "It's more exact". Anne concurred with Alex on this point because "if you had a percentage you know exactly where you stand". She worried that if two people received a merit grade you could not accurately compare their achievements because there was such a wide range of achievement within the merit level: "One could have got a high merit and one a low merit". Personally she would want to know where her particular achievement might lie on such a spectrum if she had received a merit. She was very pleased with her excellence grade.

Steve was the only one to comment on Kathy's role as both their teacher and assessor for NCEA, and he had no qualms: "I like that. You have just to be yourself and do what you can do best". He liked the immediacy with which you learned your grades for internally assessed standards "because with external exams you don't actually know your mark until the next year. I want to know if at the end of the year I've actually passed so I don't have to do Fifth Form again".

4.3.4.10 A successful outcome

With Kathy's support and teamwork all members of the group reached base camp and beyond. Prior experience, and superior abilities and skills gave some climbers an edge over others, but practice proved the means for success for the less confident and experienced. Kathy encouraged climbers to set their sights high, and aim for the top. Despite some early complacency all climbers improved their performances, much to Kathy's satisfaction as mountain guide.

4.4 Summary

This chapter has presented the findings from the two case studies in narrative style to illustrate the '*what*', '*why*' and '*how*' of student learning in daily classroom life under NCEA. The metaphorical backdrops for the case studies (a drama production and a mountain climbing expedition) help to convey a sense of learning to perform well under pressure, which was a common theme in both case studies. This performance requirement of assessment figured prominently in the student-experienced curriculum for the NCEA Science Achievement Standard 1.1 *Carrying out a practical investigation with direction* in both case studies. The metaphors help make other similarities between the findings of two case studies and some of the differences more evident to the reader; and to facilitate the discussion of key themes emerging from the data. The next chapter discusses these themes in relation to the research questions.

CHAPTER FIVE

CHAPTER FIVE

INTERPRETATION AND DISCUSSION OF RESEARCH FINDINGS

Overview

Before the interpretation and discussion of the research findings begin, this chapter reiterates the rationale for this study as it relates to the wider educational context and the focus of the research carried out in the two case studies. The revisiting of the research questions provides the framework for the interpretation and discussion of the findings as reported in the narratives in Chapter 4. Key themes arising from the narratives are identified and discussed, first as they relate to the research questions and then in light of the current literature. Differences and similarities between the two case studies are presented in order to expand the discussion and help establish transferability.

5.1 Introduction

This study arose from a desire to inform national curriculum redevelopment work currently underway in New Zealand in the field of science education. Several reviews of the existing research literature in science education (Carr et al., 2001, Hipkins et al., 2001) were commissioned to inform this curriculum redevelopment, and Carr et al. noted the paucity of locally based research into the operational curriculum that students experience in New Zealand. They therefore had to rely for much of their analysis for their literature review, into the effects of curricula and assessment on pedagogical approaches and education outcomes in science education, on large-scale reviews of research and meta-analyses of international and New Zealand literature.

Carr et al. commented that inclusion of case studies of classroom practice would have greatly enhanced the value and validity of their review by providing in situ, contextualised knowledge to explain and elaborate on the conclusions they had drawn from the large-scale studies. They believe direct study of the interplay of the more intricate and specific variables of each classroom environment, such as teacher expertise and student interactions, could provide further valuable insights into features that impact on teaching and student learning.

The international literature regarding trends in science education reports that students often achieve little meaningful learning, in terms of learning outcomes that reflect genuine scientific inquiry, from the practical work they commonly carry out in classrooms (Atkins & Black, 2003; Carr et al., 2001; Driver et al., 1996; Haigh, 2005; Hipkins & Booker, 2002; Hodson, 1996; Garnett & Garnett, 1995; Tytler & Swatton, 1992; Woolnough, 2001). Much of the practical work students engage in at senior secondary level reportedly focuses on recipe-style laboratory exercises and a 'control of variables', or fair testing, model of science investigation, which involves closed problem-solving (Mayer & Kumano, 1999; Watson et al., 1999), and learning outcomes that are predominantly content and skill-based (Haigh, 2005). As noted above most evidence for these trends at secondary level comes from large-scale quantitative studies. These studies use data obtained through surveys of teacher and student reports and opinions on classroom practice (Carr et al., 2000; Watson et al., 1999), and from some studies based on observations of 'typical' classroom practice (e.g., Berry et al., 1999; Rennie et al., 2001; Laws, 1996, White & Fredricksen, 1998). What is different between my study and most other studies is that relatively little of the international evidence about student inquiry learning comes from interpretive case studies of current classroom practice where the students' investigative activities can be seen in context and rich detail (Atkins & Black, 2003), and where participants have given perspectives on the student learning that results from these classroom activities. However, the study by Berry et al. (1999) into the nature of laboratory work being undertaken by students in several Australian secondary schools is an example of interpretive research that does provide such rich, contextual data. The findings from the Berry et al. study supported the trends identified in the large scale studies, showing that students engaged in largely closed investigations where learning tended to be superficial, and limited to following a given method and completing the task.

Other case studies into student investigative science that exist in the literature are often concerned with implementing a specific pedagogical intervention and determining its impact on classroom teaching and learning, rather than providing a 'window' into common classroom practice (Chin & Kayalivizhi, 2002; Haigh, 2005; Hart et al., 2000; Hipkins & Booker, 2002). In a classroom-based study in New Zealand, Hipkins and Booker (2002) report on a case study where a teacher introduces strategies to promote authentic student investigations under NCEA, but there is no analysis or account of any student learning that resulted.

The findings from the two classroom-based case studies in this research project can therefore begin to address the need for locally based research into the operational science curriculum to complement the extensive body of science education literature that has built up in recent times, and to investigate some of the important issues being identified in large-scale studies and meta-analyses (Jones & Baker, 2005). As in other countries, many influences within the New Zealand educational system are brought to bear on the curriculum students actually experience in classrooms. These influences emanate from sources, or sites (Carr et al., 2001; English, 1997; Knapp, 2002), within that system that support and promote particular interpretations of 'worthwhile' learning. Examples of such sites of influence in the New Zealand educational scene include the national curriculum policy statement Science in the New Zealand Curriculum [SiNZC] (MoE, 1993); curriculum support materials such as commercial publications; government educational support services, including provision of professional development for teachers; national qualifications such as the National Certificate of Educational Achievement (NCEA); school and community aspirations for the education of students; and teachers' beliefs and values about teaching and learning. These sites exert varying degrees of influence on the operational curriculum, and research indicates some influences like national qualifications may play a more significant role in shaping the student-experienced curriculum than others (e.g., Atkins & Black, 2003; Carr et al., 2001; Harlen & James, 1997; Harlen & Crick, 2003; Reay & Wiliam, 1999; Vogler, 2002). Cognitive, social, and language processes that are occurring within the classroom environment also impact on this student-experienced curriculum, along with decisions that students themselves make consciously about learning (Nuthall, 1997).

This study addresses the nature of the student-experienced curriculum where students are learning how to perform science investigations for a component of the NCEA qualification, known as Science Achievement Standard 1.1 *Carrying out a practical investigation with direction*. While the New Zealand-based literature gives some indications as to the nature of this particular student-experienced curriculum, most evidence is anecdotal, derived from the views of teachers or personnel involved in providing teacher development for the implementation of national curricula and qualifications. This study gives insights into the student-experienced curriculum based on direct classrooms observations and/or the experiences and opinions of students themselves. The findings are of value to the international arena because they reveal the educational reality for students who are experiencing a classroom curriculum assessed by a national, standards-based qualification, the NCEA Science Achievement Standard 1.1 *Carrying out a practical investigation with direction.*

The findings of this study are discussed in relation to the main research purpose presented in Chapter 1, which is to:

> Investigate the nature of the student-experienced curriculum for the NCEA Science Achievement Standard 1.1 *Carrying out a practical investigation with direction.*

This purpose is addressed by answering three questions concerned with:

- *What* science are New Zealand science students learning in NCEA classroom programmes for Science Achievement Standard 1.1 *Carrying out a practical investigation in science with direction*?
- *Why* and *how* are New Zealand science students learning the science they learn in NCEA classroom programmes for Science Achievement Standard 1.1?
- What match is there between the intended (those of the SiNZC and the teacher) and operational science curricula for New Zealand science students?

As outlined in Section 3.3.3 on data analysis, I had looked to constructivist, sociocultural and linguistic theories of teaching and learning for guidance to distinguish the *what, why* and *how* aspects of student learning in both cases studies. The key findings for these aspects of the student-experienced curricula are now summarised, and discussed in light of the relevant literature on science curricula, pedagogy, assessment and qualifications, and student learning. I attempt to draw out similarities and differences between the two cases studies as this discussion evolves.

5.2 The Science Students were Learning

5.2.1 The Science content

The evidence of student learning gathered from classroom observations of their actions, interviews and assessment information found that the science students were acquiring in the teaching and learning programmes of both case studies was very similar in most respects, and linked to a particular form of scientific investigation – fair testing. Students' classroom experiences focused on investigating cause-and-effect relationships between physical phenomena, and their thinking and learning revolved around how to plan, carry out and report the findings of fair tests into these relationships. These findings indicated that in both classroom settings the students progressively learned concepts and skills about science investigations that reflected those broadly defined in the NCEA Science Achievement Standard 1.1 *Carrying out a practical investigation with direction*, and those specifically required to complete the generic planning and reporting template for assessment tasks provided by NZQA and the MoE. This knowledge and skills were reinforced by published texts (Cooper et al., 2002; Hannay et al., 2002) that were used by students.

This interpretation of the findings is based on the detailed match between the nature of the student learning evident in the findings and the content of various assessment tasks, particularly the assessment schedules, provided by the MoE and commercial text for exemplary and practice purposes in classrooms. Students’ oral and written language, for example, showed increasing use of terminology associated with fair testing and understanding of the protocols that were prominent in these NCEA materials, such as: the independent variable and changing it systematically; the dependent variable and repetition of measurement; processing data and line graphs and lines of ‘best fit’; and interpretations of data as findings related to the purpose of the investigation. The findings suggest it was the assessment schedules of these tasks that effectively prescribed the concepts, vocabulary, skills and procedural knowledge students were gaining during the investigation exercises.

Strong indications of the influence of assessment tasks and their schedules on students’ learning also came from examination of the fair test plans that they produced as part of their investigations, and observation of their actions in implementing these plans. In exam-style sessions in the summative assessments, all students were able to individually produce plans that varied in quality from feasible (could be workable but lacks a few details) to workable (could be followed independently without further clarification). The minority of students who produced workable plans had identified and controlled key variables, and described means of obtaining, recording and processing relevant, accurate and reliable data if the plan was carried out as written.

It is important to note that there was little evidence of open-ended planning and investigation when students were required to do full investigations. All plans closely adhered to the experimental design inherent in the planning template used for assessment tasks and teachers had given considerable direction and support to the students about the content of these plans prior to planning sessions. This teacher direction provided the particular scientific relationships to be investigated and relevant experimental skills and techniques, even to the point of identifying the independent and dependent variables to be investigated in one of the case studies. The students thus went into planning sessions, even for the summative task, well informed about the procedural knowledge needed for that investigation. However, there were differences between the two case studies in the depth of understanding and level of experience with the background science concepts that students brought to the summative planning sessions, which impacted on their abilities to link their findings with science concepts. For instance, in Case Study A the students lack of familiarity with the background science involving pendulums meant they were unable to make sensible links.

As in the formative investigations, the students in both case studies worked in groups for the practical summative sessions with the result that all students had the opportunity to access workable plans. Both case study groups produced data from their experimental work, which allowed them to continue the processing, interpreting and reporting aspects of the investigation. However, since I was unable to directly observe the summative practical session in one case study I cannot compare the students' performance from the two case studies in this aspect of the investigations. In the case study where I was present during the practical work I did observe that the ability of students to collaborate successfully in the refinement and performance of their plan in the summative investigation faltered at times. These students had gained an appreciation of the need for trialling to gauge the workability of their plan and experimental methods from their formative experiences, but the group focus on decisions to do with technical details occurred at the expense of decisions to do with method. Hasty last minute decisions about procedure ultimately proved to be costly. In addition, some critical logistical points were overlooked like task delegation, and as a consequence planned decisions were not always adhered to in the summative assessment. These student actions suggest that their understanding of some of the finer points of experimental design, such as repetition and appreciation of the depth of forward planning needed were superficial, and their level of experience and expertise with the technical components of this experimental context (the pendulum) limited. Despite these ‘procedural hiccups’ the students knew what data would be sufficient to allow them to accomplish the rest of the task, and they made a pragmatic decision to ‘cook their results’. In this sense they did not achieve their teacher’s intended curriculum of ‘good science’ by fabricating results, but they did demonstrate an understanding of how to effectively meet assessment specifications. This action gave each group member access to a set of seemingly valid and reliable data, which they subsequently recorded and processed in their individual written reports.

In their written reports all students in the case studies individually recorded and processed their data sufficiently accurately enough to identify a relationship between the independent and dependent variables. Most students in Case Study A made correct data processing decisions in choosing to use line graphs, but several had trouble correctly drawing a line of best fit. Key table and graphing features that were missing from most their formative scripts were addressed in their summative scripts. In contrast, in Case Study B some of the finer details of data recording protocols for tables were missing, and all students graphed their data using bar graphs, instead of the more appropriate line graphs for identifying cause-effect relationships between two variables.

Students in both case studies were able by the close of the summative assessment to draw conclusions based on their findings and related to the purpose of the investigation. Generally speaking, however, few of the students demonstrated the capacity to fully interpret and explain their results by linking their findings to existing science concepts, and while some students were attempting to evaluate the robustness of their findings even a very able student only managed a superficial critique of his methodology. In one case study lack of familiarity with the background science in the summative task hampered students in their ability to link their findings with theory, to the extent that even an able student who had some success with this aspect in a formative task could not succeed in the science context of the summative task.

5.2.2 Links to the literature

The evidence emerging from this present interpretive study into a student-experienced curriculum demonstrates strong parallels between *what* students were learning about scientific investigations in these New Zealand classrooms and those learning trends identified from the international literature in the introduction to this chapter. In the case studies *what* students came to perceive and experience as scientific investigation was the single, linear and unproblematic methodology of fair testing. Nevertheless, within a narrow context of fair testing, the students did manifest many of those physical and mental skills generally agreed upon in the literature as the ‘scientific process skills’ (Harlen, 1999). For example, they were able to produce appropriate scientific investigations, obtain relevant information, interpret evidence in terms of the question addressed in the inquiry and communicate the investigation process (Harlen, 1999). However, with the support of planning templates, exemplar assessment materials and teacher direction the standard simply required students to follow a set of rules and procedures which they learned in practice assessments rather than coming up with original solutions to experimental design. The planning templates students used in these case studies served as blueprints, in effect restricting *what* students were learning about planning to following a formula, just as Roberts and Gott (2002) observed happening for students performing investigations under similar assessment conditions for Science Attainment Target 1 of the National Curriculum in Britain. Consequently in my case studies, students' ability to identify investigatable questions was not evident, simply because the nature of the teacher direction and the structure of the planning template did not require students to identify ways in which their scientific understanding could be expanded via investigation. The purpose of the investigation was a 'given', and students' only task was to craft a question specifying the cause and effect relationship they were investigating – in fact, in one of the case studies the teacher's direction even extended to identifying the independent and dependent variables for students in her lead in comments to the assessment. Students were not participating in authentic open-ended investigations, where they had the responsibility for determining the purpose of the investigation and the question to be investigated, as 'real' scientists would (Hodson, 1992; Reid & Yang, 2002).

Some authors comment that in terms of *what* students learn about carrying out authentic scientific activity, students tend not to learn to take account of scientific theory in planning their investigations and interpreting their results (e.g., Atkin & Black, 2003; English & Wood, 1997; Hodson, 1992). That claim is not fully supported by the findings of this study because students for the most part did take account of some scientific theory in the performance of their investigations. For example, they called on their prior learning of scientific concepts, skills and procedural knowledge related to fair testing and the scientific context to complete their planning template and conduct their investigations. The more able students also made some valid links between their findings and their scientific existing understanding, but in problem-solving situations that Reid and Yang (2002) would define as more closed in nature than open since teachers gave the students substantial guidance with the goals of the investigation, the scientific background and the procedures to be used. As a result of these student investigations there was little evidence that these activities were generating conceptual change for them as learners (English & Wood, 1997). The practical work appeared to be serving more illustrative purposes by reinforcing rather than expanding students' existing scientific understanding. In this sense, while students were practising skills and gaining experience with procedural aspects of fair testing students were not engaging in activity that reflected authentic science investigation.

In summary, for most students in these case studies *what* they learned about scientific investigation was confined to applying a 'set of rules' about fair testing, to illustrate and confirm scientific concepts covered in the instructional part of their classroom programme and to meet assessment requirements. The use of templates and exemplars in the teaching and learning programme produced the 'seen exam' phenomenon described by Roberts and Gott (2002), providing the required protocols for assessment success, and not requiring students to demonstrate the sort of tacit, intuitive knowledge in their science investigative abilities that comes with wide experience and understanding (Hodson 1992), such as creative thinking in experimental design. Students' learning was characterised by lower to middle order thinking (Bloom et al., 1956; Anderson & Krathwohl, 2001), with only a few able to display some higher order critical thinking skills. The nature of the learning for most students tended to be focused, routine, rote and superficial: rather than divergent, varied, inventive and deep-seated.

5.3 *Why* and *How* Students were Learning this Science

This section considers the reasons *why* Year 11 students in these case studies learned concepts, skills and procedural knowledge to do with fair testing, and *how* this student learning occurred. The two *why* and *how* aspects of the student-experienced curriculum are considered together in this section because while at times in analysing and interpreting the findings it was a straight forward matter to distinguish between the two, at other times it was often quite difficult. I came to the view that in some instances this exercise was becoming quite pedantic and served no useful purpose in revealing the nature of the student-experienced curriculum. For example, some explanations of *how* students learned, like interacting with peers, could be equally justified as reasons *why* they learned and vice versa. For this reason I decided to treat the *why* and *how* of students learned as consequences of:

- the content of their teachers' intended curricula
- the pedagogical approaches and techniques that their teachers used, and
- the learning strategies that students employed.

Each of these contributing factors to the student learning is discussed in turn for both case studies, including factors that influenced and shaped the specific nature of each part, and the circumstances that led to particular student learning. Similarities and differences in the learning environment of the two classrooms and the experiences of participants are drawn out through the discussion.

5.3.1 Content of the teachers' intended curriculum

At first glance, since the findings in both case studies show a close match between the teachers' intended curricula and that experienced by students, the most obvious reason *why* students in these case studies learned about fair testing in science and assessment procedures for NCEA, is that the teachers made the decision to deliver this particular content in the teaching and learning programmes. These decisions meant the teachers' instructional intentions focused on concepts, skills and procedural knowledge to do with investigating cause-effect relationships between variables and meeting the assessment requirements of the achievement standard. Clearly if certain content was not included by teachers in their programmes, then the likelihood of this 'extra' knowledge being accessed by students via classroom teaching was limited.

Close examination of the findings shows that these teacher decisions about lesson content were influenced most directly by their respective school departmental guidelines for delivering Science Achievement Standard 1.1 *Carrying out a practical investigation with direction*. These guidelines were, in turn, based on materials (planning templates, and exemplar assessment tasks and schedules) provided by the MoE and NZQA to support teaching and learning programmes for the NCEA Science

Achievement Standard 1.1 *Carrying out a practical investigation with direction.* Teachers' classroom curriculum planning decisions were not directly influenced by the specific requirements of the SiNZC as stated in the document, but more by interpretations of that national policy by NZQA and school science departments. Similar interpretations to NZQA of the SiNZC requirements were also promoted by other sites of influence, including teacher professional development providers and support agencies, and publishers of textbooks. The many similarities between the students' experienced curricula in the case studies, the teachers' intended curricula, and the science content promoted by assessment support materials and providers with the learning measured by Science Achievement Standard 1.1 *Carrying out a practical investigation with direction* provides a strong indication of *why* the learning these Year 11 students achieved in the case studies focused on fair test investigations and assessment procedures. It emerged that the science assessed by the Science Achievement Standard 1.1 *Carrying out a practical investigation with direction* was chosen by teachers, curriculum support agencies and textbook publishers as the basis for the content of the curricula they delivered. The achievement standard also had a direct influence on *how* that content was learned, as revealed in later sections of this chapter.

In Case Study A, students were also exposed to the notions of 'good science' as opposed to 'school science' in their learning, and the reason *why* stemmed from the teacher's own knowledge base and beliefs about the nature of scientific investigation. Her personal experience of scientific research gave her insights into authentic investigations that were reflected in some aspects of her content choice for the classroom curriculum; namely, her inclusion of the 'good science' notion. This added emphasis resulted in students in Case Study A receiving more in-depth treatment of certain procedural knowledge, such as trialing, reliability and systematic errors than students in the other case study, which may explain the extra emphasis these students placed on repetition and evaluation of their practical work compared with students in the other case study. These emphases on aspects considered ‘good science’ by the teacher in Case Study A were not evident in the support materials, such as the assessment tasks and text, and did not receive the same prominence in the curriculum students experienced in Case Study B. Possible reasons for this notion of ‘good science’ not featuring in the programme of Case Study B could be that the teacher did not have personal experience with scientific research to the same extent as her counterpart in Case Study A, nor the same beliefs about the nature of scientific investigation.

5.3.2 Classroom pedagogies

Other explanations for *why* and *how* students learned the fair testing and assessment procedures identified in the case studies relate to the learning environments created by the pedagogical practices of their classroom teachers and supported and fostered by the school culture. The pedagogies used to deliver the curriculum content to students need to be considered a reason for the learning, since students perceived that many of the teaching strategies their teachers used helped them to learn about investigating scientifically.

Despite the two teachers having their own distinctive philosophies about teaching and learning and other factors contributing to subtle difference in the classroom learning environments, like the prevailing school culture, classroom protocols, and teacher experience, the pedagogical approaches teachers used for the teaching of practical investigation in these cases studies were similar in many ways. There were many commonalities in the pedagogical strategies teachers employed, and in those their students thought best helped learning. The following paragraphs look at the impact of these pedagogies on the nature of the classroom learning environments, including students' responses and actions to explain in part *why* and *how* students were learning.

The pedagogical approaches the teachers employed were essentially didactic in nature, and typically included instructional sessions where the teaching was purposeful. These sessions included: guided reading and discussion work; questioning to gauge students' prior knowledge and level of understanding; demonstrations; explanation of exemplars and templates to guide students in fair test investigations; and feedback and feedforward related to the performance criteria of assessment tasks for the achievement standard, as outlined in the assessment schedules. In these sessions students learned by recalling prior knowledge, processing and linking new information with existing ideas, testing ideas in response to teachers' questions, questioning teachers to clarify understanding, and imitating the actions of the teachers. Both teachers combined instruction with sessions where students carried out exercises set from text/workbooks including practical work, and performed group investigations for practice purposes.

A significant feature of the teachers' pedagogy in the case studies that underpinned *why* and *how* students learned was the explicit sharing of learning goals and success criteria via the achievement standard, planning templates and assessment schedules with the students. This sharing of the intended learning gave students exemplars of the learning they were expected to achieve and models they could try to emulate as they applied their developing scientific knowledge and skills in the investigative work. The teachers also elicited ongoing feedback information from students about their learning progress in relation to the exemplary materials; and frequently provided feedback and feedforward information to students usually in terms of achievement as required by the assessment schedules of investigative tasks and how to improve that achievement. Students used this information to gauge their level of achievement and understanding of the science and assessment requirements, and to identify gaps or misunderstandings. More importantly, these students acted on this feedback information given during class activities, including the mock assessment, and took remedial steps that improved their achievement in the summative assessment. Teachers' feedback and feedforward comments were both planned and unplanned, and given to students in a range of circumstances from one-on-one, small group and whole class situations. Students therefore had some choice in how they could access teacher guidance.

Typically teachers used assessment information gained during classroom activities to gauge the level of understanding of content and the pace of content delivery during lessons. In the lesson following the mock assessments both teachers used assessment information gleaned from that practice activity to carry out remedial teaching. This teaching was conducted in the whole class forum and directed at common gaps in students' ability to address generic aspects of the assessment schedules that teachers had identified in the mock assessment. The teaching content consisted of tips, reminders and instances of how these identified omissions could be rectified in a given scientific context, such as factors affecting reaction rates. Of significance, there was no opportunity for students to utilise this assessment information for more learning of science in another classroom investigation, because the next investigation they attempted was the final summative assessment.

In these two classroom environments, the teachers' directions clearly prompted students' learning actions, but students did not always respond in the same way. For example, in teacher-led discussions only some students in both case studies responded to teachers' whole-class questions, or questioned teachers in return. Other students tended not to participate in the whole-class forum, but were more forthcoming in small group situations. Most students attempted tasks their teacher set and largely completed them, however, the strategy of one student was to survey all the set work, and choose only to do certain questions were the answers were not immediately evident to him - he perceived the others as unnecessary and an unprofitable use of his time. Another student used similar tactics, not to use her time more profitably, but to catch up on lost ground due to absences and some off-task behaviour in class. She finished the majority of the set tasks, much of it in her own time out of class. A student with English difficulties attempted set work in class, but regularly left gaps in her written records, noting areas of difficulty in her diary to discuss and resolve later with her tutor. Another student did attempt the set exercises, relying heavily on peer support for guidance and clarification in class, but he was frequently off-task. He often appeared not to make much progress, but out of class he used his time to finish much of the work the teacher set.

There were also instances when students in one case study responded differently from students in the other case study when given similar teaching instructions. For example, when asked to work together in groups for investigations the students in Case Study A interacted in group planning and performing, but relationships sometimes became strained when the group did not operate as a team. Frustration set in when the group's initial failure to delegate jobs in the mock assessment proved time wasting. These lapses in collegiality sometimes proved to be detrimental to learning when the group failed to deliver carefully deliberated consensus decisions in their plan, such as their inability to complete the full set of runs in their pendulum experiment. In contrast, students in Case Study B tended to look to the expertise of a particular student within the group and accepted his/her planning and performing decisions as appropriate with minimal dissension. The recognition by these students that more able students produced superior quality outcomes and that they could learn by emulating the work of these able students explained *why* and *how* some students in Case Study B learned.

There were other instances where some types of student response to teacher direction were more evident in one case study than the other. For example, in Case Study A it was more common to observe students learning by imitating skills and procedures that the teacher demonstrated and repeatedly emphasised. Students would mimic and practise these skills in follow-up sessions. However, in Case Study B students spent more time learning by processing and interpreting information to answer questions in text and student workbooks.

The teaching strategies that had most impact on *why* and *how* students were learning are now summarised, since these strategies are the ones students themselves reported as most beneficial to their learning. Students from both case studies considered the most effective ways their teacher helped them learn about doing investigations included:

- providing them with opportunities to do experiments and practice assessments in groups. The hands-on experience of scientific process skills in collaboration with peers increased their confidence in performing investigations, as they grew more aware of what was expected of them.
- the teachers' direct instruction, especially sessions where their teacher taught them beforehand the skills and background science they needed to do each investigation and the steps involved in writing up their investigations properly. Students valued their teacher's knowledge and experience highly, especially their ability to perform demonstrations that showed them how to do practical tasks, and to provide notes, definitions and explanations at a level they could understand.
- the planning template and assessment schedules that the teacher made available for use during these investigations. Students felt the constant practice with write-ups of investigations, using the templates and assessment schedules, was instrumental in their learning. These learning tools gave students the added confidence of knowing what was required to do things correctly and precisely, that is, the structure of fair tests, the terminology to use and the detail required in their write-ups
- the feedback they received from the teacher and fellow students after assessments.

Students in Case Study A also rated the teacher asking them questions, and answering students' questions as helpful, along with giving tips about various aspects of the investigative process. They acknowledged the 'spotting' technique' (i.e., targeting questions at individuals) as particularly effective because it kept them focused on what was being learned, and helped avoid being unable to supply the answer when it was their turn to field a teacher's question. Students also considered the use of the student workbook in the course valuable for the learning outcomes and the practice exercises it provided. Students in Case Study B had some specific individual preferences in terms of teaching strategies, such as the provision of notes for revision purposes and vocabulary lists or summary lists at the end of topics, but all rated highly being given interesting work to do by the teacher.

Students and teachers in both case studies were ambivalent about the value of peer assessment in promoting and facilitating learning. In Case Study A, while some students mentioned frustration and confusion as outcomes of marking other students' work in interviews, they all acknowledged in the survey that peer assessment helped their learning, especially receiving feedback from others. Their teacher recognised the potential for learning in the strategy, but worried about the accuracy of the marking since the student assessors were not picking up some common errors. In Case Study B peer assessment did not feature as an effective means of promoting their learning in students' eyes. One student expressed the view that while it was useful to talk through an assessment schedule with a peer, it was better to have the teacher mark the work because she had more scientific knowledge and expertise and was more likely to give correct information.

Finally two teaching decisions that emerged as important determinants of *why* and *how* students achieved successfully or not related to the science context in which the investigations were carried, particularly for the summative assessment, and the timing of the teaching and learning programme in the school calendar. The teaching decisions regarding the science context of the student investigations affected students' abilities to make links between their prior knowledge and their new experiences. In Case Study B the teaching decision to set both the formative and summative investigations in familiar science contexts can explain *why* and *how* students' learning was facilitated because they had a relative wealth of prior understanding and experience of the phenomena being investigated. This pre-existing knowledge gave students in Case Study B the opportunity to make meaningful links with their new experiences more readily than students in Case Study A, where the background science in their summative assessment was unfamiliar to the students and they had had little exposure to the phenomenon being investigated.

The timing of these teaching and learning programmes in the school year needs comment in terms of the opportunities for students to experience investigations in a wide variety of contexts, and to develop the tacit, intuitive knowledge required for effective investigating in science. Both school departments in the study required their teachers to deliver their programmes for Science Achievement Standard 1.1 *Carrying out a practical investigation with direction* early in the year, and final assessment of the standard was over well before mid-year in one case study and by mid-year in the other. There therefore appeared to be limited opportunity for students to consolidate and improve this learning in different science contexts. It would be interesting to know if teachers gave students further opportunity to plan and carry out investigations later in the programme – departmental guidelines gave no indication that this would or should occur.

5.3.3 Students' learning strategies and decisions

Other indications *why* students learned the science and assessment procedures covered in the teaching programme, and explanations of *how* students learned that content, can be attributed to the manner in which students engaged with the teaching in order to learn, and the reasons behind decisions students made about learning.

An obvious interpretation of the findings *why* students learned particular content relates to perceptions students had about what is valuable or important to learn. Most students in the case studies were strongly motivated to learn because they had come to appreciate and understand the 'high stakes' nature of the NCEA qualification, and had accepted achievement in an external qualification as important personal goals. These aspirations grew from the high expectations placed on students to succeed in NCEA by their teachers, schools, and families. Students felt that success in Science Achievement Standard 1.1 *Carrying out a practical investigation with direction* meant acquiring particular science content and understanding of assessment procedures, and they learned to recognise what was required of them in order to achieve the standard at particular levels of attainment. Students in Case Study A acknowledged that getting credits for NCEA, and achieving merit and excellent grades was a strong inducement to learn the necessary content for success. Initially they were content with simply gaining credits at credit level, but these modest goals were soon replaced by the desire for higher grades as their learning progressed and their confidence improved, and an element of competition with peers crept in. Students in Case Study B seemed similarly motivated, although the ESOL student was very keen to do well from the outset. In both case studies there was significant encouragement from classroom teachers for students to raise their sights and aim for merit and excellence grades in the NCEA qualification. This challenge to 'aim high' appears to have provided students with further motivation to learn.

As already discussed many of the strategies individuals appeared to use when learning, that accounted for *how* they were learning, were responses to teacher directions, such as recalling prior knowledge and making links to answer a teacher's question, processing and applying information to meet the requirements of set tasks in texts, workbook and assessment activities, or revising for assessment. Some responses involved students interacting with their teacher in individual, group and whole class situations; or collaborating with peers in pairs or small groups students to establish better understanding and application of concepts, skills and problem solving procedures. Other strategies students used to learn were not responses to teacher direction but were self-initiated, like questioning of the teacher and/or peers, experimenting, consulting text and using electronic dictionaries. There were times when students made conscious decisions about what they thought was worth doing and learning, and what was not, and acted accordingly. These actions were both reasons for *why* students learned and instances of *how* they learned

Some student learning actions were directly observable like imitating a teacher's actions after a demonstration, while others resulted in evidence of learning like modelling their work on exemplar material and seeking, receiving and acting on feedback information. Students utilised exemplar materials, particularly assessment schedules, as explicit depictions of the required learning outcomes, and subsequently modelled their own performances on these exemplars. They used these assessment schedules to interpret feedback and feedforward comments received from their teacher or peers. It is not obvious from the findings whether students gleaned information about their learning progress themselves from self-assessment techniques, or if they acted on any self-assessment information. Similarly it was not possible to determine the extent and nature of other possible student actions, such as the processing and assimilating of study notes and definitions provided by the teacher, since they were less observable and measurable. However, students rated these materials highly in terms of assisting their learning, so it is likely they were interacting actively with the materials in some way.

A noteworthy finding common to both case studies about *how* students learn, was students' practice of choosing strategies to best fit their learning needs at given times. Generally speaking, when the students were familiar and confident with concepts, skills and procedures they chose to learn independently using strategies that best worked for them: reading over notes; working through problems and questions from text; and revising using summaries or word lists. However, when unsure, the students frequently turned to their peers to question, consult or confirm ideas. They often chose this tactic ahead of approaching the teacher for assistance. This consultation with peers was particularly prevalent in the group work during investigations.

Students in both case studies cited working in pair or groups as playing a key role in *why* and *how* they learned to carry out investigations– the sharing of knowledge and expertise to problem solve, clarify misconceptions and confirm and consolidate understanding were outcomes of these social behaviours that they believed contributed to their learning. This perception that they learned from one another was supported by the classroom observations that I made of their interactions and behaviours as an unobtrusive researcher. It also emerged from the findings that many students placed high value on being able to work with their peers when and if they needed to. This preference was often born of convenience because peers were more readily accessible, or from feeling more comfortable exposing gaps in understanding to peers rather than the teacher. Fellow students had the added advantage of providing explanations that were considered more comprehensible in linguistic terms than the teacher's version. Such interaction with peers allowed students the means to engage in learning at a pace and in ways that suited their immediate learning needs. If students sensed gaps in their own understanding, or needed reassurance that their thinking and ideas were appropriate, the immediacy of their peers for support and guidance was invaluable. This type of interaction often helped students to build bridges in their understanding. Other interactions often prompted and encouraged students to draw on their prior knowledge and experiences and engage in new learning as a collaborative effort – this form of peer teaching and learning became most evident when the students performed investigations in groups. These interactive opportunities allowed students to monitor and scaffold their own learning to some extent.

However, on some occasions these peer interactions compromised learning. The interactions, for example, that led to the 'cooking of results' incident in Case Study A resulted in outcomes that were contrary to the notion of 'good science' that their teacher held, and the ideals of integrity and honesty. Students themselves considered that working with friends was not always helpful to their learning and could have a downside. Several students commented that off-task behaviour by other students could be distracting, or further confusion could result after consultation with peers over difficulties simply because their peers had a similar lack of scientific understanding or even misunderstandings. Interestingly, there were instances of strategies involving peer interactions that to the outside observer engaged students in more critical thinking, and hence facilitated opportunities for deeper levels of understanding. Students did not always have the same perception of these methods. As noted earlier peer assessment, for example, was seen more by students as frustrating learning rather than promoting it. On balance though, the students believed the ability to work with their peers was beneficial to learning, because it helped to build their understanding. Many students regarded asking the teacher for assistance with difficulties as a good fallback strategy if the help they received from peers did not improve their understanding or capability to perform a task.

As a participant classroom observer I witnessed student collaborations impacting on the nature of their learning. For example, the highly able student in Case Study B regularly worked in step with his neighbour. Together, as a unit independent from both the teachers and the rest of the class, they achieved learning at a pace and level that suited their particular abilities and needs. Group collaboration generally assisted all students' learning, but outcomes for individuals, or for groups, in similar learning situations were not necessarily the same. The group planning and performance of the investigations for the summative assessment in the case studies is an instance when these differences occurred. In Case Study B the participants did not work together in a single group but as members of disparate groups. The learning outcomes for any group, as a collective, were easy to detect for certain aspects of the investigative process, such as planning and gathering data, but not so for individual members. The groups had been deliberately selected to work together by the teacher, such that each group contained at least one member who had demonstrated advanced capabilities in their planning of the investigation. Thus, when the groups came together to negotiate an agreed plan, this collaboration usually resulted in the members performing a workable plan. The students acknowledged that the negotiated plans their particular groups operated on closely resembled those of the member students who had produced workable plans as individuals. It would be difficult to judge on the basis of these negotiated group plans whether each individual then possessed the capability to design a workable plan for the specific question in the given context. However, the collaborative process gave group members further opportunity to access the additional ideas required to make the group plan workable, and could be a possible reason *why* some students learned to plan, and contributed to *how* they learned to plan. These resulting workable plans gave all group members the potential to secure relevant and reliable data if the plan was accurately performed, and in turn the chance to process and interpret data, draw conclusions and evaluate their findings.

In Case Study A, where the participants worked together as a group for all their investigations, the consultative process in the summative assessment produced some different outcomes. A possible reason *why* students did not achieve outcomes by the 'preferred method,' by making an inappropriate procedural decision and overlooking some critical logistical points, was because they did not have a group member with planning expertise in this context. Their lack of experience and expertise with the technical components of this experimental context affected *how* they learned by limiting their ability to gather relevant data. Students' relative unfamiliarity with the background science of pendulums, and lack of procedural understanding about graphing values for the independent variable also contributed to their data gathering difficulties. It is interesting that despite these 'procedural hiccups' some group members knew what data would be sufficient to allow them to accomplish the rest of the task, and influenced others in the group to make a pragmatic decision to 'cook their results'. This action demonstrated that these particular students had a good understanding of how to effectively meet assessment specifications and why they achieved success in some aspects of the standard's requirements. Given the group's circumstances at the time this decision proved to be pragmatic, because it gave each group member a set of data, which they could in turn process and interpret to complete the assessment task as individuals. The teacher did not detect this 'massaging of results', so the students were not penalised. Had the students not taken this decision, it is possible that without a complete set of data they could have failed the standard. Peer interactions and collaborations can explain *why* and *how* students achieved these learning outcomes to do with assessment procedures.

Other instances of deliberate learning decisions by students occurred when students purposefully disengaged from learning activities for reasons that they considered valid. One student, for example, would daydream or chat if he already knew the content being taught to prevent boredom, while another only participated in hands-on practical work because he preferred and enjoyed learning this way. Several students chose not to waste time going over exercises that were repetitious and non-challenging; others chose their homework time to carry out much of their learning. One student identified language difficulties in class in her journal for later referral work with her language tutor, rather than seek help in class.

These findings indicate *why* and *how* students achieved certain learning by illustrating that students make conscious decisions themselves about the nature of their learning outcomes and how they went about achieving those particular outcomes, and can have preferences which determine learning.

5.3.4 Links with the literature

In considering *why* and *how* the Year 11 students in these two case studies achieved their learning, the close match that was found between the teachers' intended curricula in these case studies and that experienced by their students is significant. This finding shows agreement with findings from the literature, that maintain teachers' instructional intentions have a direct bearing on *why* and *how* students learn (Atkin & Black, 2003; Hargreaves & Fullan, 1992; Lederman, 1999; Mc Gee & Penlington, 2001b; Tytler, 2003). However, the claim that classroom teachers are the only ones whose actions directly affect students' learning (Harlen & Crick, 2003) needs to be qualified in these case studies, in light of the obvious similarities between the operational curriculum occurring in classrooms, and the SiNZC interpretation promoted by the Science Achievement Standard 1.1 *Carrying out a practical investigation with direction.*

5.3.4.1 The influence of the achievement standard

English (1997) reported, from her work with New Zealand teachers who were preparing students for Level 1 unit standards, that the interpretation of the curriculum portrayed in NZQA qualifications at that time had a powerful effect on New Zealand senior classroom curricula because of the ‘high stakes’ perception of the qualification amongst teachers, students and the community. It appears this effect also holds true for the achievement standards of NCEA, and this observation is supported by similar overseas experience where high stakes testing and qualifications are also reported to drive classroom practice (McDonald & Boud, 2003; Orpwood, 2002; Preece & Skinner, 1999; Roberts & Gott, 2004; Tytler & Swatton, 1992; Wiliam, 2000; Wynne, 1999). Acknowledging the absence of any recent large-scale studies of actual classroom practice in secondary schools in science teaching in New Zealand, and the unavailability of base-line data against which change in classroom practice could be measured, Hipkins (2004) used teachers’ own perceptions of changes they have made in the delivery of their Year 11 science programmes since the introduction of NCEA to gauge the impact of the qualification on classroom curricula. In a study involving 18 teachers, the teachers’ comments from their interviews show clearly that they are adapting their classroom practice to meet NCEA requirements, including some who expressed the view that student investigations “had to be focused within the narrow, formally presented framework of the reporting schedule for this achievement standard [i.e., Science Achievement Standard 1.1 *Carrying out a practical investigation with direction*]” (p. 9). The findings from the case studies in the present research also illustrate the pervasive influence of the assessment regime underpinning the NCEA qualification on the teachers’ planning intentions, since teachers felt responsibility to help their students achieve success in the qualification and their classroom curriculum design was effectively pre-determined by decisions made at departmental level. That decisions about the content of these departmental guidelines were strongly influenced by the requirements of a national qualification, and that all teachers in the respective departments were required to follow these guidelines closely when implementing their own classroom programmes, must be taken into account when considering whose actions directly affected students' learning. Departments in both case studies made school-based decisions that impacted on the nature of the curriculum experienced by students, and it will be argued that the actions of more than just the classroom teacher had direct bearing on the students' learning.

The close similarity in each case study between the teachers' intended curricula, the departmental guidelines, and the interpretation of curriculum promoted by the NZQA in its NCEA qualification suggests that the teachers, and science departments, in these case studies were for the most part acting as conduits for the achievement of that government agency's goals. Many decisions to do with classroom practice were effectively taken out of the individual teachers' hands – judgements were made instead collectively at departmental level and were based on guidelines and recommendations from NZQA, which classroom teachers were obligated to follow under school accreditation requirements for NZQA. This departmental layer of interpretation took into account some of the key external determinants of these Year 11 classroom curricula (McGee & Penlington, 2001a), and effectively made decisions that the classroom teachers were obliged to implement in their classroom curricula. These decisions included the:

- content of the teaching and learning programme

- manner in which the teaching and learning programmes were to be delivered and assessed, with the emphasis on classroom procedures and arrangements for the practice (formative) and summative assessments and methods of moderation
- timing of the programme delivery
- adoption of the planning template as recommended by NZQA, and
- the use of exemplar assessment tasks and schedules supplied by the MoE for NCEA as the basis for teaching and assessment materials for use across all classes in the department.

However, as active members of their respective science departments, the teachers in these case studies would have had at least some role in creating this layer of curriculum interpretation and pedagogical and assessment approaches presented in the departmental guidelines. Their participation in meetings concerned with marking and moderation were likely forums for their contributions to be heard and incorporated into the departmental guideline. The strong similarities between the respective departmental guidelines and intended curricula of the teachers in both case studies also lend strong support to the contention that the NCEA qualification had an over-riding influence on the teachers' instructional intentions.

The actions of individuals from other sites of influence (English, 1997), namely those writers who created the national assessment guidelines and exemplar materials for NCEA and the published text used in the classroom programmes in these case studies, also had a direct effect on students' learning in classrooms because students interacted frequently with these materials in their daily classroom activities and home study.

5.3.4.2 Teachers' beliefs, values and pedagogies

The beliefs and values teachers held about science and the teaching and learning of science, and the pedagogical styles of both teachers, could also have had a direct bearing on *why* students learned particular content in both case studies and *how* this learning occurred. In a study of the classroom practice of 37 high school science teachers in Taiwan, Tsai (2002) noted an alignment between teachers' beliefs about science and their views about teaching and learning; these views were often associated with particular pedagogical practices. For example, a teacher with a positivist view of science would view teaching as transferring knowledge via illustrative activities and learning and assessment as replication of that knowledge. Kang and Wallace (2004) report comparable trends in their study of five high school science teachers in the USA, where they were investigating the relationship between the teachers' epistemological beliefs, teaching goals and their use of laboratory activities. I found similar relationships in my case studies where the teachers' perceptions of science, as portrayed in their comments and actions, seemed aligned more to an empiricist or logical-positivist view than to science as a human invention (Beck, 1979; Bullock & Trombley, 2000; Burns, 1994; Cohen et al., 2000; Lather, 1992). Both teachers tended to use transmissive methods of teaching, characterised by instructional sessions for the provision of content, and practical work focused on learning specific procedures for performing fair testing and meeting assessment requirements through repetitive practice. Their pedagogical styles suggested neo-behaviourist views of learning (Neyland, 1995), where learning involves a behavioural response by students. Learning, in this view, is a relatively passive process of absorption of knowledge as it is transferred from the mind of the teacher to the mind of the student (Nuthall, 1997). In this sense, both teachers would fit the description of presenters of content rather than leaders of an exploration (Black, 2003a), implying that the explicit and repeated exposure of students to the content to be learned was an explanation for *why* and *how* students learned.

Another finding could also support the view that teachers' empiricist beliefs about science influenced their views about teaching and learning (Murphy, 2003; Tsai, 2002). Students' learning in both case studies reflected the linear approach to scientific investigation as depicted in the achievement objectives of the *Developing Scientific Skills and Attitudes* strand in the SiNZC and the Science Achievement Standard 1.1 *Carrying out a practical investigation with direction*, rather than the more holistic view also expressed in the SiNZC that scientific activity and scientific knowledge are interdependent. However, it would be difficult to assert on the basis of this evidence that the pedagogies of these teachers in the case studies were a direct consequence of their beliefs about science and the teaching and learning of science, when the requirements of the departmental guidelines are taken into consideration. In both case studies the departmental guidelines left little room for teachers to vary their pedagogical content and approaches away from fair testing and the transmissive model of teaching even if they were inclined to.

A further reason for *why* students' learning focused on the experimental design of fair tests and the assessment requirements of NCEA is that both teachers placed high value on students 'succeeding' in their learning, and they equated students' learning success with high levels of achievement in qualifications. To this end, the teachers did focus in their pedagogy on equipping student with the knowledge, skills and procedures necessary for achieving the standard. Their pedagogical practice also strongly influenced *how* students learned by providing them with detailed content via exemplary material, practice with skills and procedures, and experience with a full mock or practice assessment. The pedagogy of the teachers featured many hallmarks of formative assessment including: routine gathering of information through teacher-student interactions, and to a lesser degree student-student interactions, in relation to reference levels depicting expected performance or achievement (Sadler, 1989); continuous, ongoing feedback that supported learning by indicating the actions needed to improve teaching and learning (Bell & Cowie, 2002; Black & Wiliam, 1998; Glover & Thomas, 1999); and actions based on the assessment information that led to improved learning (Harlen & Crick, 1997). The key pedagogical function of the form of ongoing assessment occurring in these case studies was to discover if the students knew the procedure of fair testing as depicted in the assessment schedules for tasks based on Achievement Standard 1.1. Teachers adopted an analytic approach (Sadler, 1989) in their assessment practice using manifest criteria that were made clear to students from the start, and students "came to understand what counted as good work through a focus on the criteria and on their exemplification" (Black, 2003, p. 8). In this respect the assessment was convergent in nature (Torrance & Pryor, 2001) and more akin to the practice of continuous summative assessment commonly found in other high stakes assessment situations (Atkins & Black, 2003; Carr et al., 2001; Harlen & James, 1997; Harlen & Crick, 2003; Reay & Wiliam, 1999). As well as assessing fair testing capability, the assessment clearly provided feedback on how well students were handling the assessment procedures of NCEA – in other words it was serving a second function of "assessment for assessment" Hipkins (2004, p. 6).

While key elements of formative assessment were evident in classroom practice in both case studies, some aspects of that practice reduced the potential of the assessment to maximise the students' learning. The extent to which teachers could utilise the formative assessment information to adapt their teaching programme to meet students' emerging learning needs (Black, 2003; Black & Wiliam, 1998; Clarke, 2001; Sadler, 1989) was restricted by the departmental guidelines, as were the opportunities for students to take the next learning steps. In both case studies, the short duration of the teaching and learning programme in terms of allocated classroom time created an 'over-crowded-curriculum' effect (Crooks, 2002a), and encouraged superficial as opposed to deep learning by 'rushing' the learning. The timing of the programme early in the year also reduced the effectiveness of the formative function of the assessment by giving students only one opportunity to experience a full investigation before they were summatively assessed. As a consequence, students were missing out on exposure to investigative activity in a variety of science contexts and the "prolonged engagement in evaluative activity" (Sadler, 1989, p. 135) that would help them gain more of that tacit or intuitive scientific knowledge that comes from experiencing investigations in many different contexts (English & Wood, 1997; Duggan & Gott, 1995; Hodson, 1992, 1995). This 'crash-course' classroom curriculum undoubtedly contributed to *why* the learning experienced by students in the case studies had a narrow focus, and *how* they learned via absorption and replication of knowledge. Deboer (2002) talks of the potential for tension when students are curtailed in their freedom to carry out authentic inquiry by the prescribed content of standard because teachers and students feel pressured to cover that particular content. The specified nature of the standard and its requirements did contribute to the narrowness of the student experienced curriculum, but it is important to note that the decisions dictating the timing and time allocated to the teaching of this investigative unit were made at school and departmental level, and they were not set requirements of the NCEA qualification or the NZQA.

Hipkins et al. (2004) report a prevailing view among some heads of departments in New Zealand secondary schools that attaining a high number of overall credits was superior to gaining excellences but with fewer overall assessment credits. It could be that decisions by the schools to provide science courses with high credit numbers, and hence overcrowded curricula, were influenced by similar perceptions of teachers that the quantity of credits gained in NCEA was a criterion by which success in national qualifications could be gauged. The decision to time the unit early in the year, it seems, was one of expediency leaving more time for teachers and students to concentrate on the externally assessed standards later in the year.

Another departmental decision impacting on student learning that was not required by NCEA, and is worth noting, concerned the science context in which assessment tasks were set. Choosing a context unfamiliar to students in one case study led to an able student struggling with his explanations of the findings when in an earlier investigation where he was conversant with the background science he had been successful in explaining his findings. This student’s inability to explain results in an unfamiliar science context supports the argument that students need a strong theoretical or conceptual background in the science context of the investigation in order to use scientific theory to make sense of their findings (English & Wood, 1997; Harlen, 1998; Hodson, 1992; Leach & Scott, 2003; Luft, 1999).

5.3.4.3 Sociocultural factors

Finally in considering *why* and *how* students learned the content identified in the two case studies it needs to be acknowledged that their learning cannot be attributed to just a cognitive exercise on their part, but also to factors in the sociocultural and affective domains (Black, 2003, Carr et al, 2003; Nuthall, 1997). It seemed that to initiate and sustain this cognitive engagement, students needed to perceive good reasons for learning (Brophy, 1999). Students cited, for example, that personal interest in the content being taught, such as their liking for experimental work, prompted their learning by arousing their curiosity. Acquisition of the NCEA Achievement Standard 1.1 *Carrying out an investigation with direction* also emerged as good cause for learning in most students eyes because they appreciated the value and importance that society placed on this particular learning (Brophy, 1999). Students too perceived the NCEA qualification as valuable and high stakes, and an important goal to achieve.

Other motivating factors were related to students' feelings of self-esteem and confidence in their ability to succeed. Much of the students' growing sense of self-efficacy, that is, their ability to use judgements on their performance to decide whether they are capable of undertaking a task successfully (Harlen & Crick,2003), can be attributed to the teachers' assessment practice which scaffolded students' learning by the sharing of assessment criteria, exemplification and feedback (Crooks, 2002; Sadler, 1989). This structured approach resulted in most students coming to understand the worth of what they were learning (Brookhart & Bronowizc, 2003), and provided them with reason for needing or wanting to do the necessary learning. Since much of the feedback was task-oriented rather than ego-oriented (Black, 2003; Black

& Wiliam, 1998; Clarke, 2001), the self-esteem of many students rose as they came to realise that improved achievement could result from effort, and was not simply down to ability. It appears these same students found the feedback information intrinsically motivating (Brookhart, 2001), perhaps because they came to believe that they could achieve success through effort (Black, 2003a; Clarke, 2001). Unlike some reports from the literature (e.g. Black, 2003a), the giving of grades for the formative assessment by peers in Case Study A did not appear to de-motivate students in their learning, although it could be argued that students were mastering the ‘rules of the game’ rather than improving their procedural understanding of investigations (Keiler & Woolnough, 2002). The fabrication of results by students in Case Study A in the summative investigation suggests they knew how ‘to play the game’. The competitive element that the giving of grades introduced, witnessed by their interest in sharing their achievements with their peers in class, did seem to spur some students on in their efforts to improve their performance. Again while the very able students in both case studies continued to attribute their success to ability rather than effort (Black, 2003a), these students did rectify gaps identified in feedback information from their formative exercises in their future performances.

Affective factors, like feelings of comfort and confidence about their learning may also in part explain the finding in the present work that students chose particular learning strategies to best fit their learning needs at given times. Such decisions illustrated that students could assume some control over their learning as they mediated the nature and pace at which it occurred (Carr et al., 2003; Nuthall, 1997). As Black (2003) found in his observations of classroom formative assessment practice in British high schools, students in these New Zealand case studies were more likely to approach a fellow student if they did not understand an explanation than interrupt the teacher. Such actions were often more convenient, since the help was at hand and immediate, and also posed less threat to some students' self-esteem. These students perceived that disclosure of personal inadequacies in a very public way, such as asking questions of a teacher in class, could expose them to the ridicule of their peers. Bell and Cowie (2001) reported a similar reluctance by some students to disclose lack of understanding in their classroom study of formative assessment practice in New Zealand secondary schools.

In this inquiry the strong preference of students to work with their peers on learning tasks lends support to sociocultural and linguistic views of *how* learning occurs (Barnett & Hodson, 2001; Leach & Scott, 2003; Nuthall, 1997), and the notion that interacting positively with peers to co-construct understanding can be a powerful aid to learning (Bishop & Glynn, 1999). Students commented on how they valued the positive reinforcement that came from interchange with their peers (Leach & Scott, 2003), and the ability of peers to provide explanations in forms of language that students used naturally in their everyday talk (Black, 2003a). For successful learning to emerge from peer work on group tasks, Black (2003a) comments that guidance from the teacher is needed so students can appreciate how to cooperate and assign responsibilities within the group. The failure of a group in the study to work together effectively, seen here in some aspects of their investigative tasks, may have been due to insufficient guidance from the teacher and experience with how to behave when tackling such tasks. Similarly, in both case studies some students' seemingly contradictory and negative comments about peer assessment may be explained by their lack of experience with the strategy and failure to appreciate its power to promote learning. Students need time and practice to develop the skills of peer assessment (Black, 2003a, Black & Wiliam, 1998; Sadler, 1989).

This section has discussed *why* and *how* students learned *what* they did in their experienced curriculum for the NCEA Science Achievement Standard 1.1 *Carrying out a practical investigation with direction*. The next section reflects on the match between intended curricula and the student-experienced curricula to consider the effectiveness of national curriculum policy implementation

5.4 The Match between the Intended Curricula and the Student-experienced Curriculum

5.4.1 The mismatch between the aims of the SiNZC and the student-experienced curricula

In interpreting and discussing the findings to the two previous research questions in sections 5.1 and 5.2 of this chapter, the close match that existed between the teachers' intended curricula and the student-experienced curricula in the two case studies has already been identified and recognised as a highly significant factor in the nature of the curriculum experienced by students in this study. So too has the finding that the content of these two forms of curricula show strong parallels with the content assessed by Science Achievement Standard 1.1 *Carrying out a practical investigation with direction* for the NCEA qualification. The Achievement Standard 1.1 as an interpretation by the New Zealand Qualifications Authority (NZQA) of the requirements of the SiNZC policy statement [MoE, 1993b], closely reflects the achievement objectives in the *Developing Scientific Skills and Attitudes* strand at Level 5/6 of the SiNZC (the age-appropriate level for Year 11 students in the New

Zealand schooling system). The intent in this section of the chapter is to discuss the mismatch that has occurred between the aims of the SiNZC and teacher intended and student-experienced curricula

Within the intended curriculum of the SiNZC, a disparity can be seen between aspects of the general aims and achievement aims of the strands, and the achievement objectives at Levels 5/6 particularly in the *Developing Scientific Skills and Attitudes* strand. The introductory sections of the curriculum statement speak of scientific investigations in broad terms, such that a wide range of scientific investigations could be embraced within the description including fair tests, surveys, systematic exploratory work, and experience of physical phenomena and research. In contrast the achievement objectives at Levels 5/6 tend to focus on experimental design, notably fair testing in the *Developing Scientific Skills and Attitudes* strand. This is further reinforced by the suggested learning experiences, most of which involve students carrying out experimental investigations. There are some exceptions in the suggested Level 6 learning experiences of the contextual strands such as identifying the presence of particular ions in various substances, interpreting seismograph recordings and various research exercises. The focus on experimental design and the fair test from the Level 5/6 achievement objectives has been adopted by the Science Achievement Standard 1.1 *Carrying out a practical investigation with direction*. Therefore students would be unable to achieve the requirements of the standard if they were to carry out an investigation such as interpreting seismographs. In fact many investigations in the contexts of the Earth Sciences Achievement Standards 1.5 and 1.6 to do with rocks and minerals and astronomy could not be appropriately assessed by Science Achievement Standard 1.1. This belies the information given in the guidelines to the standard that states the context should come from content drawn up from Level 6 of the SiNZC.

5.4.2 Links to the literature

McGee (1997) argues that the judging the effectiveness of a national curriculum development and its implementation can be gauged by the degree of match between the ‘intended curriculum’ and the ‘operational curriculum’, while Bell and Baker (1997) maintain that a curriculum and its implementation can only be judged effective if the student-experienced curriculum changes in a way that improves learning. The findings from the case studies indicate that there are close matches between the teachers’ intended curricula and that experienced by students, but there are some mismatches between these curricula and the intended curriculum as portrayed in the SiNZC that can be traced back to the national policy statement itself. The narrow interpretation of scientific investigation as fair testing that is contained in the expected learning outcomes for students at Levels 5/6 has resulted in this specific emphasis becoming a key focus of the relevant NCEA qualification component, that is, Science Achievement Standard 1.1 *Carrying out a practical investigation with direction*. Since this high stakes qualification had a dominating influence over the nature of student-experienced curriculum in these cases studies students thus gained a restricted understanding of authentic scientific investigation. This dominating influence of a high stakes qualification is not dissimilar to other overseas experiences where teachers give more attention to that content which is tested and ‘teaching to the test’ in effect governs the classroom curriculum (Harlen, 1998; Preece & Skinner, 1999; Roberts & Gott, 2004). It would appear that there is some basis to the fears that most New

Zealand students will only come to appreciate scientific activity as a simple linear process rather than a complex mix, as expressed by Hipkins and Barker (2002). Since Year 11 is the last formal schooling in science for many New Zealand students, and the Science Achievement Standard 1.1 is the only achievement standard assessing practical investigations for NCEA it is unfortunate that these students could be leaving schools with an unrealistic view of the nature of science (Carr et al., 2001).

This mismatch between the aims of two key national policies and the operational curricula may detract from the effectiveness of the curriculum implementation process (McGee, 1997). In terms of improved learning outcomes for students (Bell & Baker, 1997) that reflect the authentic practice of scientists (English & Wood, 1997; Duggan & Gott, 1995; Haigh & Hubbard, 1997; Hodson, 1995), the curriculum implementation seems to have only been partially successful. Students are gaining concepts, skills and procedural knowledge to do with fair testing, but not of the many other forms of inquiry that scientists engage in.

5.4 Summary

By describing and discussing the key findings from the case studies under the three main research questions, this study has been able to identify important features of the student-experienced curriculum for the NCEA Science Achievement Standard 1.1 *Carrying out a practical investigation with direction*. Perhaps the strongest influences on the nature of the classroom curriculum that students came to experience are the content of the standard itself and the standards-based structure of the NCEA qualification. Schools and teachers in the study looked to the high stakes NCEA qualification to guide curriculum delivery decisions in their Year 11 science classrooms, and effectively treated the content of Science Achievement Standard 1.1 as a prescription. This emphasis by schools in the case studies on the achievement standard was reflected in the teachers' predominantly didactic pedagogical approaches and assessment practices. Consequently students' learning in the case studies was purposeful and focused on the fair testing aspect of scientific inquiry and assessment procedures for NCEA.

Student learning was promoted by pedagogies that incorporated many features of formative assessment including the explicit sharing of achievement criteria, exemplification of expected outcomes and feedback in relation to these reference levels. The effectiveness of formative assessment in enabling students to develop higher order thinking skills was reduced by the limited opportunities for students to act on feedforward information in a range of science contexts. Consequently most students successfully demonstrated lower level capabilities of experimental design in the case studies, but few showed higher order critical thinking skills.

Students mediated their learning in terms of choices they made about when to engage in learning, the manner in which they engaged and the strategies they employed to learn at given times. They had preferences for particular strategies, most favouring a degree of teacher direction, hands on activities and opportunities to work with peers. They were motivated to learn by the requirement to gain a qualification, the assessment practices of their teachers and content that was of personal interest.

These findings indicate that students in New Zealand schools may not be experiencing authentic scientific inquiry as portrayed in the SiNZC, or recommended by the international science education. The reasons seem to lie in the curriculum interpretations and decisions that are made within the education system at national policy level, at sites of influence such as qualifications providers and in schools, school science departments and classrooms. The implementation of the SiNZC can be regarded as ‘unsuccessful’ in this regard.

The next chapter includes recommendations for policy makers, qualification providers, schools and teachers based on these findings and suggestions for possible research arising from this research.

CHAPTER 6

CHAPTER 6

IMPLICATIONS AND CONCLUSION

Overview

This chapter highlights the implications of the findings from this inquiry into the student-experienced curriculum for NCEA Science Achievement Standard 1.1 *Carrying out a practical investigation with direction* for national science curriculum redevelopment, the NCEA qualification, the practice of schools and their science departments, and classroom teaching, assessment and learning. Recommendations are made for actions to address issues raised in the implications. The trustworthiness of the study is considered in terms of the limitations of the findings, and suggestions are made for future research. The final section of the chapter contains concluding statements about the inquiry.

6.1 Introduction

This research was primarily interested in finding out *what* New Zealand students were learning about scientific inquiry in Year 11 of their schooling, and *why* and *how* this learning was occurring, to inform the current redevelopment of the national science curriculum policy *Science in the New Zealand curriculum* (SiNZC) and classroom practice. The study has provided a window into the operational curriculum for Year 11 students from the perspective and the views of students themselves and their teachers, and thus gives valuable insights into how national curriculum policy is manifested in New Zealand classrooms and *what*, *why* and *how* students actually learn.

Since it is hoped that applicability of the findings from this study to other situations is a key research outcome it is important at this point to alert the reader to the limitations that can exist in an interpretive study of this kind. In drawing conclusions from the findings of these two unique case studies, consideration needs to be given to how the issues of comparability and transferability, and dependability and conformability have impacted on the trustworthiness of this study. These issues will be considered in more detail later in this chapter in Section 6.6.5.

Under the present educational regime in New Zealand, the findings suggest Year 11 students working on Science Achievement Standard 1.1 gain a rather narrow view of the nature of scientific enquiry by focusing student learning on the scientific concepts, skills and procedural knowledge of fair testing. This learning appears to be happening despite the national policy goal that students should develop authentic understandings of and capabilities in science as widely recommended by international science educators, and in this regard the implementation of SiNZC appears unsuccessful. This potential for scientific inquiry to be 'misrepresented' to students seems to stem largely from the over-riding influence that the high stakes National Certificate of Educational Achievement (NCEA) qualification and the interpretation of the SiNZC promoted by the NCEA Science Achievement Standard 1.1 *Carrying out a practical investigation with direction* have on decisions determining the nature of the operational curriculum. Decisions at school and science departmental level to effectively treat Science Achievement Standard 1.1, along with the support materials provided by the New Zealand Qualifications Authority (NZQA) and the Ministry of Education (MoE), as a prescription for their Year 11 investigative science ultimately left little room for classroom teachers to make substantial contributions to their teaching focus and approach in classrooms, apart from their contribution as members of the department to departmental curriculum decisions. Teachers chose to deliver the curriculum using didactic pedagogical and assessment approaches, and directive roles in students' investigations. Students consequently engaged in inquiry work that was essentially closed in nature, rather than open, and guided carefully by teachers. The structure of NCEA, including its mode of assessment, did give teachers a format for utilising many key features of convergent formative assessment in their pedagogy to successfully scaffold many aspects of student learning. However, in practice, the full benefits of formative assessment on student learning appear limited by restricted opportunities for students to act on feedback information and to develop mature peer and self-assessment capabilities. Consequently much student learning took on a formulaic and rote flavour and few students demonstrated higher order creative and critical thinking skills that A.S 1.1 did promote such as the ability to evaluate findings and link them to existing scientific concepts. Despite the didactic pedagogical approaches students often played a mediating role, at times consciously choosing when and how to engage from a range of personally preferred learning strategies. These decisions were influenced by affective factors such as self esteem, self confidence, motivation and competition.

This chapter now considers the implications arising from these findings for the redevelopment of the New Zealand science curriculum, and for countries and educational communities who may be considering or undergoing similar developments in their science curricula. Suggestions are made for curriculum policy makers and personnel responsible for the implementation and assessment of the science curriculum in New Zealand, and for future research.

6.2 Implications for National Curriculum Policy-Makers

A key concern for curriculum policy-makers emerging from the science education literature about scientific inquiry is that students should come to understand the epistemological basis of science and how that knowledge is constructed through scientific inquiry and argumentation (Sandoval, 2005). If that understanding, which many science educators believe comes from students engaging in authentic scientific inquiry and constructing knowledge as scientists do (e.g., Atkins & Black, 2003; Duggan & Gott, 1995; Hodson, 1992) is to be promoted in New Zealand classrooms then the message must come through to the science teaching community in the national curriculum policy. The degree of mismatch between the stated goals of the current SiNZC and the student-experienced curriculum as identified in this study, and the nature of those learning outcomes for students, suggest that perhaps this message is not being received by those involved in the delivery of the curriculum in schools and classrooms. One of the reasons contributing to this appears to be some ambiguity in the SiNZC statement itself and the mixed messages it is sending about the nature of student learning outcomes. Thus as this curriculum policy was implemented, the potential for a number of different interpretations was possible as participants in various sites of influence within the education community were able to put interpretations on the policy that placed greater emphasis on certain aspects of the curriculum than others, perhaps reflecting their particular perspectives, values and beliefs. When one site of influence, that is, the NZQA and its qualification NCEA, was perceived or considered by the educational community to be more influential in terms of what students should achieve in their learning, this site's interpretation of national policy prevailed when decisions about curriculum content and delivery in classrooms were made by teachers and departments. Consequently student learning became focused on experimental design associated with fair testing and with standards-based assessment procedures.

If science curriculum policy-makers wish to promote student learning *of* and *about* science via scientific inquiry, then the obvious implication for curriculum design is that this goal and subsequent aims and achievement objectives must convey the same message. This message needs to be stated in clear terms to reduce the likelihood of divergent interpretations from various sites of influence within the New Zealand education system as it appears in this study to have happened with the SiNZC. A draft statement for the redeveloped New Zealand science curriculum, which has been recently released (MoE, 2005a), shows signs of some encouraging shifts in emphasis regarding the nature of science and inquiry from that portrayed in the SiNZC. The two SiNZC integrated strands *Making sense of the nature of science* and *Developing scientific skills and attitudes* have been subsumed into one over-riding strand called *Developing scientific competencies*. This new strand has four Achievement Aims entitled:

- understanding about Science
- investigating in Science
- communicating in science, and
- participating and Contributing

All these aims seem to present science more as a form of human activity with its own social practices that results in a particular kind of human knowledge that is both durable and tentative. The complexity of scientific investigation is more visible in the *Investigating in Science* draft achievement aim statement, which clearly states the varied approaches students are to experience in their investigative work:

> Students will carry out science investigations using a variety of approaches: classifying and identifying; pattern seeking; exploring; investigating models; fair testing; making things or developing systems. (MoE, 2005a, p. 4)

This intent is reiterated in part in the Level 5/6 achievement objective derived from this aim:

> Students will develop and carry out investigations that use a variety of approaches. Variables will be considered, and logical and justifiable conclusions drawn. (MoE, 2005a, p. 5)

It could be argued that inclusion of the phrase "variables will be considered" might infer an emphasis on fair testing for many readers, so it would be important that any back up statements, exemplars and materials provided by the MoE to support this curriculum objective, exemplify the role of variables in other approaches to investigation. These provisions should encourage other sites of influence within the New Zealand education system, such as published text, professional development providers and teachers, to interpret the national policy as intended. However, the experience in the UK, as reported by Laws (1996), would suggest that despite refinements to curriculum aims and objectives as described above the classroom reality of student investigations in New Zealand schools is likely to remain little changed.

6.3 Implications for NCEA

The redevelopment of the New Zealand science curriculum will have implications for the NCEA qualification since an aim of NZQA is that NCEA embrace the New Zealand Curriculum (NZC) and support classroom programmes (Lee & Lee, 2002). The intent to reflect national curriculum policy in NCEA is also evident in the stated links to the relevant parts of the NZC in each of its achievement standards (see Appendices A and Q). If the final version of the redeveloped national curriculum maintains a requirement that students experience a variety of approaches to scientific inquiry in its *Investigating in Science* achievement aim and corresponding Level 5/6 achievement objective, then Science Achievement Standard 1.1 will require substantial redrafting. One solution may be that Science Achievement Standard 1.1 have a series of sub-sections within it, each section representing a group of investigation types that can more readily described in generic terms because they share many common features. For example, classifying and identifying investigations may sit comfortably with pattern seeking investigations, or with exploring. Students would be assessed in the whole set of sub-sections in order to achieve the credits currently ascribed to Science Achievement Standard 1.1. A necessary flow on effect of any redrafting of Science Achievement Standard 1.1 in this manner would be the need for exemplary materials and text to support teachers in creating optimal learning environments and students in achieving the required learning outcomes. Issues of workload and assessment overload would also need to be considered.

Already in the intervening period since the collection of data for this study and the write-up of the thesis, NZQA has made some modifications to Science Achievement

Standard 1.1 *Carrying out a practical investigation with direction.* In October, 2005 the standard was re-registered with a number of changes (see Appendix R). These changes on the surface seem to introduce more recognition of the complexity of scientific investigation into the standard and give more latitude for teachers to offer students some variety in their approaches to scientific investigation. The revised standard also provides more specific detail about what constitutes 'quality' in a scientific investigation. The achievement criteria are more generic than those in the previous form of the standard, and some former aspects of the accompanying explanatory notes have been given increased emphasis, while some have been dropped and new features introduced. For example:

- greater specificity is provided about what constitutes a directed investigation
- the terms *practical investigation* and *quality practical investigation* are introduced and defined in detail, reflecting the content of the modified achievement criteria. The terms *workable* and *feasible* to describe plans are dropped
- the terms *sample* and *collection of data* are introduced alongside the terms *independent* and *independent variable* respectively in the definition of a practical investigation, and *sampling* and *bias* as possible factors to consider in data gathering in the description of a quality practical investigation. The inclusion of these terms potentially enables students to use approaches to investigation other than fair testing, but because *sampling* and *bias* can have close connotations with fair testing it is possible that fair testing may still prevail in classroom practice unless appropriate exemplary support materials and text are accessible to professional development providers, teachers and students.

- validity of method, reliability of data and science ideas are specified as requirements to consider where relevant when evaluating the investigation

These changes signal more acknowledgement of the nature of scientific inquiry in NCEA assessment procedures for Science Achievement Standard 1.1, and possibly greater opportunity for students to experience authentic scientific investigations and develop higher order thinking skills. An overview of exemplary material now present on the MoE website for Achievement Standard 1.1 reveals one assessment task (see Appendix R) linked to the new version of the standard. This assessment resource is based on a pattern-seeking investigation. The resource includes a planning and reporting template and assessment schedule similar in format to the fair testing versions, but with terms relevant to pattern seeking and the new requirements of the standard.

A final comment concerns the title of Science Achievement Standard 1.1 itself, and the phrase ‘with direction’. This in a sense automatically implies that ‘control’ of experimental design is not in the hands of students but rather with the teachers. It could be argued that the need for such a degree of teacher direction at this stage of students’ schooling is unnecessary and undesirable given the aim of *authenticity* in investigative learning and the previous learning about scientific investigation that students have experienced in the *Developing scientific skills and attitudes* strand as they worked through Levels 1-5 of the SiNZC before reaching Year 11. If students are learning about the nature of science and scientific inquiry and participating in authentic investigative learning episodes in their earlier schooling as intended in the SiNZC, then by Year 11 most students should be capable of attempting some inquiry independent of their teacher and achieving success. If this is the situation, then the need for the phrase ‘with direction’ in the title of the standard disappears. However, the findings from this study suggest that students were not gaining much experience of authentic open-ended inquiry in their science programmes prior to Year 11, so perhaps the inclusion of ‘with direction’ in the standard has some justification if these findings reflect the current circumstances for most Year 11 students in New Zealand. It would also be difficult to imagine teachers not assuming some degree of direction over students’ investigations given the high stakes nature of NCEA.

6.4 Implications for Schools and Science Departments

The schools and science departments in this study played a significant role in classroom curriculum design because these groups made decisions that essentially dictated the content of classroom programmes, their duration, the timing and methods of assessment, and consequently the pedagogical approaches teachers used to deliver lessons in class. These decisions in effect prevented students engaging in a range of investigative approaches and restricted opportunities for students to develop deep understanding of the scientific concepts, skills and procedural knowledge needed to do investigations with confidence and flair. Departmental decisions that allocated more classroom time to student investigative work in a range of science contexts and summative assessment later in the year could be means of increasing students’ opportunities to benefit from the pedagogical practices that teachers were using which promoted effective learning, such as formative assessment. Courses offering fewer credits may release time for more investigative teaching and learning. A greater range of experiences over a longer time period could allow teachers to help students develop the higher order creative and critical thinking skills and tacit, intuitive knowledge that is characteristic of successful learners (and scientists). Schools and departments may well consider providing teachers with the ability to make summative decisions based on evidence gathered during the programme as they interact with students and observe them investigating in a variety of contexts. However, such practice would require extensive professional development and dialogue to ensure validity, reliability, fairness and moderation issues are addressed.

Since the schools and departments in this study looked mainly to NCEA and its achievement standards for guidance in programme design, any redevelopment of the standards which may result from national curriculum change is likely to require the schools and departments to rethink and re-evaluate their science programmes. If authentic scientific investigation is the focus of any modified achievement standard then these schools may need to give due consideration to some of the points raised above.

6.5 Implications for Classroom Teachers and their Students

As described in earlier sections of this chapter and the preceding chapter, the teachers' opportunities to design teaching and learning programmes to meet students' specific learning needs and interests were curtailed to a large extent by the requirement that teachers adhere to the departmental guidelines for classroom delivery of Science Achievement Standard 1.1. Should changes to SiNZC, the achievement standard and school and departmental guidelines proceed as indicated in the sections above for the schools in this study, then there would likely be changes in pedagogical and assessment practice and student learning since teachers in these schools are obliged to follow departmental guidelines. To promote more effective student learning of authentic scientific inquiry the findings of this study suggest that these changes to departmental guidelines and classroom programmes be considered:

- adjustments to the teaching and learning content and assessment practice, with emphasis on exposing students to more divergent forms of investigation over a wider range of science contexts and over a longer teaching period before summative assessment decisions are made by classroom teachers.
- greater autonomy for teachers in the design, delivery and assessment of their classroom curricula to support the use of a wider range of pedagogical approaches and strategies, to strengthen formative assessment practice and encourage students to take a more active role in their learning. The findings from this study suggest that by placing greater pedagogical focus on developing students' peer and self-assessment capabilities their motivation to learn and quality of learning are enhanced.
- more opportunity for students to work with their peers and teacher, in open rather than closed investigations, and in science contexts where they have had substantial experience with the background science and choice in the topic they are investigating. In these investigations the teacher can assume a less directive and more facilitating role while students accept more responsibility for monitoring their own learning and acting on feedback information.

As Hipkins (2004) points out the positive benefits of NCEA on student learning that may come from new emphases in classroom curriculum, pedagogy and assessment will only eventuate if teachers are given relevant professional support. If curriculum and qualification reforms carry the potential to 'open up' the classroom curriculum to wider interpretations of scientific inquiry, new pedagogies and assessment for learning, then this study indicates science departments and teachers need professional development support primarily in the areas of authentic science inquiry, formative assessment and the formative-summative divide. This support needs to be ongoing and come from government agencies, schools, heads of department and professional colleagues.

There are number of suggestions for secondary teachers' classroom practice arising from the findings of this study that could increase the authenticity of student learning about scientific investigation and the level of their thinking skills, regardless of any changes that might occur at national policy level and in the NCEA Science Achievement Standard 1.1 and within the restrictions on classroom practice existing school and departmental guidelines may impose. These suggestions embrace not only the Year 11 science programmes but also those in the junior science area of secondary schools. Suggestions include:

- the strengthening of teachers' formative assessment practice to enhance students' learning capabilities by giving students less direct information about the detail that is missing from their learning and more opportunity to act on feedforward information generated by the teacher, peers and increasingly by students themselves. Feedforward information should be in the form of actions students need to take to address gaps in their learning and so promote students looking to the standards and exemplary materials themselves to gauge what the outcomes of those action s will look like in the particular context, rather than relying on teachers describing the specific outcomes they are to reproduce. Such formative assessment practice encourages students to be less reliant on teacher direction and more self-motivated and independent in their learning. It also supports the development of students' critical thinking skills, especially evaluative thinking. Such capabilities are not built overnight and need to be carefully scaffolded, so students would benefit by experiencing these aspects of formative assessment as integral components of their junior science programmes. Regular use of explicit criterion-based formative assessment, and peer and self-assessment techniques in junior science programmes should give students more confidence and awareness of the worth of these learning strategies.

- classroom teachers assessing the student investigation for summative assessment decisions as part of the in-class teaching and learning programme. While issues of moderation can arise, the opportunities for further on-going student learning are worth considering given the additional feedback and feedforward information classroom teachers can give individual students. Since their classroom teachers are present during the students' summative investigation and can make first-hand observations of students' actions, summative assessment evidence can take on a formative function.
- giving students explicit teaching in how to work effectively in groups. The findings from the research literature and the classroom studies in this work point to the learning benefits that students can gain when co-constructing understanding with their peers. However, students need to learn strategies to work productively in groups like thinking ahead and establishing agreed goals and targets, delegating roles and tasks within the group and regularly evaluating progress. The learning of these skills needs careful scaffolding by teachers, as with other aspects of investigative work, if students are going to acquire them as part of long-term learning. Making deliberate decisions about

student membership in groups to ensure each group contains one or more students with high levels of ability and skills is worth considering too, since the findings of this study indicate that groups are more likely to produce plans that are workable and result in relevant data collection. However, such decisions do not necessarily guarantee all students develop planning skills; they may merely continue to mask some students' inabilities to plan.

- teachers giving students prior exposure in their junior science programmes to a range of investigative approaches in genuine open-ended investigations where the science contexts are of relevance and interest to students should also have follow-on benefits for student learning in Year11 science and beyond. Such exposure should give students greater awareness of the nature of authentic scientific inquiry, more opportunity to think creatively, and lay a stronger foundation for developing critical thinking skills. Hopefully students would gain more grounding in the necessary concepts, skills and procedural knowledge required to carry out scientific investigation, not just in fair testing but in other investigative approaches as well. Such exposure could also increase teachers' pedagogical content knowledge of scientific investigation in the relative 'safety' of a teaching and learning environment not dominated by a high stakes qualification.
- changes to the format of the template that necessitate students providing an account of relevant science concepts as an introduction to the report may assist students in making later links between their findings and the background science. If students are required to carry out an investigation that is summatively assessed then prior investigative work in the same science context may also assist students with making these important linkages.

6.6 Limitations

To facilitate a holistic, interpretive investigation of events in context, the case study approach was used in this study because it has the potential to provide a more complete picture of educational practice than perhaps other modes of research. By revealing the reality of classroom life and all its idiosyncrasies from the perspective of students and their teachers, this study sought to gain a deeper understanding of classroom events and the science curriculum students were experiencing than had been previously depicted in the research literature. By probing deeply and analysing intensively the many phenomena that comprised each of the two case studies, I believe this study has generated findings that a reader could use to inform his/her/their own educational aims and practice

To contribute to the trustworthiness of my research process I paid particular attention to strategies that would maximize the quality of data gathering and processing within the constraints of my study. For instance, pragmatic reasons meant decisions had to be made about the size and composition of the research groups that potentially detracted from their representative and typicality qualities (i.e., small size of groups and their limited representation of the student population), but I believe the decision to take a two-case study approach in this study helped to mitigate the impact of these limiting sampling factors and promote transferability. The ability to compare and contrast findings from two unique case studies allowed similarities and differences to be drawn, so increasing the chances for more robust results and wider applicability.

Triangulation, of both methods of data collection and sources of data, was another strategy used to promote the dependability, confirmability and credibility of the study by reducing the likelihood of researcher bias and producing sufficient wealth of evidence to allow a high degree of convergence, despite some unforeseen difficulties obtaining field data in one of the case studies. Observation was prolonged and extensive enough to establish the dependability of the data, helped by my detailed auditing of the inquiry process and respondent validation of the raw data. These actions also endorsed the confirmability of the data, although this aspect would have been strengthened had I used respondent validation processes at the time of data collation and categorising. Time restraints and the practicalities of physically involving the teachers and students in these processes prevented me from utilising this form of respondent validation.

Finally it is hoped the rich descriptions of the case studies using narrative style has allowed the participants voices to be heard, and the analytic categories of *what*, *why* and *how* have enhanced the comparability and transferability of the research findings by giving readers greater opportunities make meaningful comparisons with their own situations.

6.7 Suggestions for Future Research

This research is among some early studies (e.g., Hipkins 2004; Hipkins & Booker, 2002; Hipkins et al., 2004) into the nature of science learning occurring for New Zealand students under the SiNZC, which has been national science curriculum policy for over 12 years, and the recently introduced standards-based assessment regime of the NCEA qualification. It is to be hoped that the findings from this research have provided some insights into what Year 11 New Zealand students are learning about scientific investigations and the nature of the operational curricula they are experiencing in classrooms. The implementation of the redeveloped national science curriculum in 2007 provides fertile ground for further research, particularly for evaluating the impact of that national policy on other sites of influence within the New Zealand education system such as the NCEA qualification, on classroom curricula and most importantly, the quality of student learning. The call from recent reviews of the international science education literature (Hipkins et al., 2001; Jones & Baker, 2005) for local classroom-based research in New Zealand, particularly case studies (Carr et al., 2001), to complement the extensive body of research that currently exists provides clear direction for the form that research needs to take.

To inform and evaluate the implementation of this redeveloped curriculum in terms of the benefits for student learning and achievement in scientific inquiry the following suggestions are made for future research:

- longitudinal case studies to monitor the curriculum implementation process and determine the impact on classroom curriculum, pedagogy, assessment and student learning outcomes. Research questions could be along similar lines to those questions underpinning this study, with particular interest in the interpretations and decisions made by sites of influence such as the NCEA, school and departmental management and classroom teachers which were found to have such a strong impact on the nature of the student-experienced curriculum in this study.

- case studies to evaluate the science curriculum in relation to student learning of scientific investigation after the implementation process is completed.
- case studies similar to this study that focus on particular aspects of the operational curriculum that were not so visible in the findings of this study, such as teachers' content knowledge and pedagogical content knowledge of scientific inquiry and students' roles and capabilities in peer and self assessment.
- action research to determine the impact of various interventions on the operational curricula and student learning. Such interventions might be: teacher professional development into the nature of scientific inquiry and appropriate pedagogies and assessment practices, or the processes of formative assessment, peer assessment, and self assessment; the use of support materials like exemplars and text; or the trialling of new approaches to teaching and learning as signalled by the research literature like co-construction of knowledge by teachers and students.

6.8 Conclusion

This study sought to find out the nature of the student-experienced curriculum in the New Zealand context as Year 11 students learn about scientific inquiry for the national qualification NCEA from the perspectives of participants in the operational or classroom curriculum. The purpose was to inform redevelopment work on the existing national science curriculum SiNZC by gaining insights into: the relationships between the intent of the SiNZC in the area of scientific investigation and that actually experienced by students, the nature of students' learning outcomes and reasons underpinning that learning; and ways of achieving improved learning outcomes for students. Little was known about the nature of this student-experienced curriculum in the research literature, particularly from the recipients' point of view. The study was exploratory and interpretive because it wanted to give a rich and detailed account of classroom life and student learning as it was actually happening through the eyes of those participants who were experiencing these events.

By examining *what* students were learning about science investigations, and *why* and *how* this learning occurred the research found that the student-experienced curriculum was focused on a narrow view of scientific inquiry as fair testing, and on acquiring assessment techniques. This discrepancy between the intent of the SiNZC expressed in its aims and the curriculum experienced by students arose because of mixed messages about scientific inquiry learning within the policy statement itself and the strong influence the interpretation of the SiNZC by the NCEA qualification was having on decisions affecting classroom curricula in schools. This qualification was considered high stakes by the schools and teachers involved in the study and consequently, as Black (2001, 2003a) observes, assessment for qualifications drives the senior school and classroom programmes. Decisions were made at school and departmental level, which reflected the importance school communities and professional staff placed on their students achieving success in this qualification, and this directly impacted on the content of classroom curricula and the methods teachers used to deliver that content.

The NCEA interpretation of the SiNZC, in the form of Science Achievement Standard 1.1 *Carrying out an investigation with direction* and the supporting materials, and departmental decisions determining time allocation and timing of the science investigation programme in classes influenced the didactic pedagogical approaches teachers chose to use, and the strategies used by students to learn. The structure of NCEA and the standards-based mode of assessment promoted formative assessment practice, and teachers employed many features of convergent formative assessment. However, relatively short teaching and learning programmes before summative decisions were made restricted students ability to act on formative assessment information to improve their learning. Consequently, student learning tended to focus on procedures and there was little evidence of the higher order thinking skills linked to creativity, evaluating and self-monitoring of learning.

The sway that the NCEA interpretation of scientific investigation had on curriculum design and delivery decisions made by schools, departments and classroom lends support to the view that moving from policy document to the operational curriculum in classrooms is not a straightforward process (Atkin & Black, 2003; McGee & Penlington, 2001). Knowing the literal translation that the NCEA site of influence made of portions of the SiNZC relevant to its purpose of assessment for a qualification should alert policy-makers to the importance of conveying a clear message about student learning outcomes. Introducing more flexibility into the Science Achievement Standards assessing students' understanding and capabilities in scientific investigation, and support materials, should facilitate improved student learning outcomes in terms of authentic scientific inquiry and greater teacher autonomy in designing teaching and learning programmes to meet students' learning needs and interests. Awareness that school-based decisions that focus too much on meeting administrative, logistical and moderation requirements of high stakes qualifications like NCEA can have detrimental effects on pedagogy and student learning, may hopefully prompt schools to re-evaluate the wisdom of these decisions. Finally the views and insights that students have given in this study, about the teaching and learning they experienced and the role they play in these processes, should provide useful information for teachers to reflect on as they evaluate the effectiveness of their pedagogical and assessment strategies in helping students to achieve quality learning in scientific inquiry.

LIST OF APPENDICES

Appendix A

Science Achievement Standard 1.1 *Carrying out a practical investigation with direction*, as at April 2004

Achievement Standard

Subject Reference		Science 1.1			
Title		Carry out a practical science investigation with direction			
Level	1	**Credits**	4	**Assessment**	Internal
Subfield	Science				
Domain	Science – Core				
Registration date	21 October 2003	**Date version published**	21 October 2003		

This achievement standard involves carrying out a practical investigation, with direction, by planning the investigation, collecting and processing data, and interpreting and reporting the findings.

Achievement Criteria

Achievement	Achievement with Merit	Achievement with Excellence
• Develop a plan that identifies some key variables.	• Develop a feasible plan.	• Develop a workable plan.
• Collect, record and process information appropriate to the investigation.	• Collect, record and process information appropriate to the investigation.	• Collect, record and process sufficient information appropriate to the investigation.
• Present a report with interpretations relating to the investigation.	• Present a report with interpretations and a conclusion linked to the purpose of the investigation.	• Present a report with interpretations and a conclusion linked to the purpose of the investigation. Provide a comprehensive evaluation or discussion of the investigation.

Explanatory Notes

1 This achievement standard is derived from *Science in the New Zealand Curriculum,* Learning Media, Ministry of Education, 1993, ‘Developing Scientific Skills and Attitudes’, p. 42–51; and *Pūtaiao i roto i te Marautanga o Aotearoa*, Learning Media, Ministry of Education, 1996, ‘Ngā Pūkenga me Ngā Waiaro ki te Pūtaiao’, p. 70–85.

2 Procedures outlined in *Safety and Science: a Guidance Manual for New Zealand Schools*, Learning Media, Ministry of Education, 2000, should be followed. Investigations should comply with the Animal Welfare Act 1999, as outlined in *Caring for Animals: a Guide for Teachers, Early Childhood Educators, and Students,* Learning Media, Ministry of Education, 1999.

3 An *investigation* is an activity covering the complete process from planning to reporting and will involve the student in the collection of primary data.

If a student enters for assessment against AS90186, Science 1.1, as well as any of: AS90156, Agriculture and Horticulture 1.1, AS90161, Biology 1.1, AS90169, Chemistry 1.1, or AS90180, Physics 1.1, the investigations must be in different 'subject' areas. For example, if a student is being assessed against AS90161, Biology 1.1, and is also being assessed against AS90186, Science 1.1, then the major emphasis of their investigation for AS90186, Science 1.1, cannot be biological.

4 Investigations should be based on situations arising from content drawn from up to science/pūtaiao curriculum Level 6. Possible contexts are given in the curriculum documents.

5 The investigation will be directed. Student procedural instructions for the investigation will be specified in writing, and templates or suitable formats for planning and reporting will be provided.

6 *Planning:*

- The plan will contain the *purpose* of the investigation. This may include an aim, testable question, prediction or hypothesis based on a scientific idea.
- The student should be provided with the opportunity to undertake some form of trialling or checking of the plan so it can be adapted if required.
- A *workable plan* includes a valid range for key variables and details of how they will be measured. The influences of other variables have been taken into account and, if necessary, methods for their control are stated. A scientific method to collect data is described and shows consideration of factors, such as sampling, bias, sources of error and sufficiency of data. Some checks to ensure the plan is workable have been made.
- A *feasible plan* is one that could be workable, but lacks detail.

7 *Collection, recording and processing of data:*

- The plan is followed and data, appropriate to the investigation, are *collected and recorded* in a table or other systematic way.
- *Data processing* would usually involve calculations (eg averaging and/or graphing) to establish a relevant pattern or trend.

8 *Interpretation and reporting:*

- *Interpretations* of the processed data that relate to the purpose of the investigation are given.
- The *report* follows the format clearly specified in written guidelines by the assessor and would usually include the following sections:
  - ***plan, including the purpose of the investigation and final method used***
- recorded data
- processed data, showing links to the recorded data
- interpretations and a conclusion, including a generalised statement linking the findings of the investigation with the purpose of the investigation
- evaluation or discussion, which may include limitations of the investigation, other variables that had not been foreseen, difficulties in measuring, difficulties in the use of equipment, limitations of the findings and impact on the outcome and, where relevant, suggested solutions or pathways for further investigation and links to science concepts or ideas.

Quality Assurance

1 Providers and Industry Training Organisations must be accredited by the Qualifications Authority before they can register credits from assessment against achievement standards.

2 Accredited providers and Industry Training Organisations assessing against achievement standards must engage with the moderation system that applies to those achievement standards.

Accreditation and Moderation Action Plan (AMAP) reference 0226

Appendix B

"Bubble Trouble" – Internal Assessment Resource Reference Number: Sci/1/1 A Version 5

Appendix C

Interview Schedules

Proposed Interview Schedules

<u>Schedule for teacher interview 1</u>

1. Tell me about your teaching background in science

Prompts:

- qualifications for teaching
- teaching experience
- professional development undertaken recently
- current position

2. Tell me about your philosophy of teaching and learning in science

Prompts:

- values and beliefs about teaching and learning
- teaching approaches used
- role of assessment
- influences on your thinking
- advantages of your approach
- barriers and constraints on your teaching approach

3. How do you develop and deliver your practical investigatory skills in Year 11 science programmes for NCEA Science AS 1.1?

Prompts:

- basis on which planning is done, and the extent and nature of that planning
- methods of delivery to students (integrated?, stand alone unit? etc)
- teaching techniques and approaches used
- nature of student involvement
- methods of monitoring learning progress
- the basis on which you judge the programme's effectiveness
- successes/challenges

Schedule for teacher interview 2

1. Are there any points you would like to comment on or expand upon from the transcript of your first interview?

2. Have you experienced any changes in your beliefs/values /philosophy since our last interview as a result of experiences since then?

3. In the NCEA training sessions did you do any training on the teaching and assessing of investigatory science for A.S. 1.1? If so, what form did it take?

4. How do you think your students' learning is progressing?

Prompts:

- what have they learned/not learned?
- how do you know?
- is it what you expected? Are they meeting expectations
- what brought about the learning - strategies, activities, resources etc?
- what were the barriers to learning?
- on reflection is there anything you would have done differently? Any surprises?

5. Tell me about the next phase of the investigatory skills programme

Prompts:

- will you continue with your programme as originally planned?
- what learning do you hope students will achieve in this next stage?
- how will you know?

Schedule for teacher interview 3

1 What science did students learn in this investigatory skills area?
Prompts:
- specific skills, techniques, concepts
- how do you know what they learned?
- was it what you expected?

2 How did they learn this science?
Prompts:
- methods, techniques students used to learn and their relative effectiveness
- teaching strategies that promoted/hindered this learning

3 Recommendations for future teaching and learning programmes in the investigatory skills
Prompts:
- science content
- teaching approach – techniques, resources, timing, assessment.
- nature of student participation
- other issues

<u>Schedule for student interview 1</u>

1 Do you enjoy science?
Prompts:
- like/dislikes
- reasons

1. Do you learn well in science?
Prompts:
- how do you know?
- why do you think you are successful/unsuccessful at learning in science?

2. Tell me how your teacher usually starts a new topic of work in science
Prompts:
- pre-test, revision, brainstorm, a challenge or problem to solve

3. Talk about the things that help/hinder your learning
Prompts:
- relationship with the teacher
- ability of the teacher – strategies
- learning skills like mind maps, summaries, revision
- classmates
- resources like computers
- family
- personal interest

4. What are your feelings about NCEA?
Prompts:
- negative/positive?
- reasons

Schedule for student interview 2

1. Are there any points you would like to comment on or expand upon from the transcript of your first interview?

2. What science have you been learning?

Prompts:
- ideas, skills, knowledge, procedures
- how do you know?
- what progress have you made in terms of what you have to know
- how are you feeling about your learning? – pleased/unhappy?

3. How are you doing your learning?

Prompts:
- types of learning activities and which were the effective ones?
- teacher's assistance or other students
- personal strategies

4. What might make you learn better in this area?

Prompts:
- factors related to time, teaching style, resources, study skills, incentives/rewards, parents etc

Schedule for Interview 3

What have you learned about science investigatory skills this year?

- key ideas, skills, techniques
- how do you know?
- are you satisfied with your learning?
- how useful is this knowledge?

What helped you to learn best?

- teacher, teaching strategies, learning activities, resources, classmates, parents etc

Looking back at the investigatory skills area, what would have improved your learning?

- things done differently
- more/less of a particular thing

Appendix D

Letter to Principal

Introductory Letter to the Board

Dear Principal

I am currently carrying out research work for my doctorate degree at the University of Waikato. My interest lies in the nature of the science that students are learning for the National Certificate of Educational Achievement (NCEA) at Year 11. I am seeking your approval to work in your school, and to approach your students and teachers to invite their involvement in my study. This letter briefly explains what my study involves and what you may need to consider before agreeing to participate.

In my research I am particularly interested in what investigatory skills students are learning in science at this level and how this learning occurs. To gather my data in your school I would carry out one case study of a Year 11 NCEA science class, involving audiotaped interviews with the teacher and students, and the observing and audiotaping of lessons. I would also need to gather information from departmental schemes, teacher planning and assessment and students' written work including any assessments. This may involve photocopying or photographing written material in some instances.

The data gathering would happen in two phases over the year, one early in the year and one later, each phase lasting 1-2 weeks (5-10 lessons). In each phase I will do three interviews (30-60 minutes duration) with the teacher before, during and after the set of classroom observations. Similarly I'll hold three interviews with a group of selected students from the class. In total this involves teacher and students in up to six interviews and on occasion informal follow up conversations to clarify some ideas. The teachers and students will have the opportunity to see transcribed notes of their respective audiotapes to ensure they are accurate records of his/her responses and actions.

I will look at the gathered data to try and develop a global view of students' learning experiences. Any reports of the research findings will thus present broad themes only and the identities of the teacher, students and school are carefully protected. Where selected data from transcribed material is used to support the summary of themes, I will use pseudonyms to again prevent identification. The findings will be presented as part of my doctoral thesis, at seminars and conferences, and published in research journals to help others involved in science education to understand the issues.

In my professional life I work as an evaluator for the Education Review Office and this places me in a position that could raise potential ethical issues. I am committed to taking every possible step to prevent harm to participants that may result from my position, including the undertaking not to be a member of future review teams in schools and classrooms of teachers and students involved in the research.

If you would like to know more, or meet with me to discuss the project before making any kind of decision, please feel free to contact me. I am most happy to elaborate on any points or discuss any concerns. My contact details are:

Home phone number 07 856 4592 e-mail Hume@hnpl.net

Work phone number 07 8381898 e-mail anne.hume@ero.govt.nz

If you agree to your school involvement and feel happy with this information please sign the consent section below and post it back to me. Once I have received your consent I will telephone you to arrange a meeting time with appropriate senior staff (for example, deputy principal and head of department) to carry out the next stage.

I look forward to hearing from you.

Regards

Appendix E

Letter to Class

Appendix F

Letter to Teacher

Introductory Letter to Participating Teachers

Dear Teacher

I am currently carrying out research work for my doctorate degree in education at the University of Waikato. My interest lies in the nature of the science that students are learning for the National Certificate of Educational Achievement (NCEA) at Year 11. I am inviting you to participate in this research, as your views are important in helping to understand the issues involved. This letter briefly explains what my study involves and what you may need to consider before agreeing to participate.

In my research I am particularly interested in what investigatory skills students are learning in science at this level and how this learning occurs. To gather my data in your school I would carry out one case study of a Year 11 NCEA science class, involving audiotaped interviews with the teacher and students, and the observing and audiotaping of lessons. I would also need to gather information from departmental schemes, teacher planning and assessment and students' written work including any assessments. This may involve photocopying or photographing written material in some instances.

The data gathering would happen in two phases over the year, one early in the year and one later, each phase lasting 1-2 weeks (5-10 lessons). In each phase I will do three interviews (30-60 minutes duration) with the teacher before, during and after the set of classroom observations. Similarly I'll hold three interviews with a group of selected students from the class. In total this involves yourself and students in up to six interviews and on occasion informal follow up conversations to clarify some ideas. You and your students will have the opportunity to see transcribed notes of your respective audiotapes to ensure they are accurate records of your responses and actions.

I will look at the gathered data to try and develop a global view of students' learning experiences. Any reports of the research findings will thus present broad themes only and the identities of yourself, your students and school are carefully protected. Where selected data from transcribed material is used to support the summary of themes, I will use pseudonyms to again prevent identification. The findings will be presented as part of my doctoral thesis, at seminars and conferences, and published in research journals to help others involved in science education to understand the issues.

In my professional life I work as an evaluator for the Education Review Office and this places me in a position that could raise potential ethical issues. I am committed to taking every possible step to prevent harm to my research participants that may result from my position, including the undertaking not to be a member of future review teams in schools and classrooms of teachers and students involved in the research.

If you would like to know more, or meet with me to discuss the project before making any kind of decision, please feel free to contact me. I am most happy to elaborate on any points or discuss any concerns. My contact details are:

Home phone number	07 856 4592	e-mail	Hume@hnpl.net
Work phone number	07 8381898	e-mail	anne.hume@ero.govt.nz

If you agree to your involvement and feel happy with this information please sign the consent section below and post it back to me. Once I have received your consent I will telephone you to arrange a meeting time to carry out the next stage.

I look forward to hearing from you.

Regards

I have read the proposal in this letter and am happy to give my consent

Signed ***Date***

Appendix G

Letter to Student Participants

Introductory Letter to Students

Dear Student

I am currently carrying out research work for my doctorate degree in education at the University of Waikato. My interest lies in the kind of science that students are learning for the National Certificate of Educational Achievement (NCEA) at Year 11. I am inviting you to take part in this research, because your views are very important in helping to understand what is involved. This letter briefly explains what my study involves and what you may need to consider before agreeing to join in.

In my research I am particularly interested in what investigatory skills students are learning in science at Year 11 and how this learning occurs. To gather my data in your school I would carry out a case study of your Year 11 NCEA science class, involving audiotaped interviews with the teacher and students, and the observing and audiotaping of lessons. I would also need to gather information from your written work including any assessments, and this may involve some photocopying or photographing of examples of your work.

The data gathering would happen in two stages over the year, one early in the year and one later, each stage lasting 1-2 weeks (5-10 lessons). In each stage I will hold three interviews with a group of selected students from the class. In total this involves those students in up to six interviews and on occasion some follow up conversations to check on some ideas. You will have the opportunity to see transcribed notes of your respective audiotapes to ensure they are accurate records of what you said and did.

I will look at the gathered data to try and develop a global picture of your learning experiences. Any reports of the research findings will present broad themes only and the identities of yourself, your teachers and school are carefully protected. Where selected data from the tapes is used to illustrate the themes, I will use pseudonyms (made-up names) to again prevent identification. The findings of my research will be presented as part of my doctoral thesis, at seminars and conferences, and published in research journals to help others involved in science education to understand the issues.

In my professional life I work as an evaluator for the Education Review Office and this places me in a position that could raise potential ethical issues. I am committed to taking every possible step to prevent harm to my research participants that may result from my position, including the undertaking not to be a member of future review teams in schools and classrooms of teachers and students involved in the research.

If you would like to know more, or meet with me to discuss the project before making any kind of decision, please feel free to contact me. I am most happy to elaborate on any points or discuss any concerns. My contact details are:

Home phone number 07 856 4592 e-mail Hume@hnpl.net
Work phone number 07 8381898 e-mail anne.hume@ero.govt.nz

If you agree to your involvement and feel happy with this information please sign the consent section below and post it back to me. Once I have received your consent I will communicate with you and your teacher to arrange a meeting time to carry out the next stage.

I look forward to hearing from you.

Regards

I have read the proposal in this letter and am happy to give my consent

Signed Date

Appendix H

Letter to Parents

Information Letter to parents/caregivers

Dear parent/care-giver

I am currently carrying out doctoral work at the University of Waikato and have invited your son/daughter to be part of the research being conducted in 2004. As a courtesy to you I am writing this letter to explain the study and what it involves.

The study is centred on the science learning of students in Year 11 and involves students and their science teachers. The views of students and their teachers are very important in trying to understand the nature of the science they are learning and how they learn that science. In my research I am particularly interested in what investigatory skills students are learning in science at this level and how this learning occurs. To gather my data in your school I am carrying out a case study of a Year 11 NCEA science class, involving audiotaped interviews with the teacher and students, and the observing and audiotaping of lessons. I will also gather information from students' written work including any assessments and this may involve the photocopying or photographing of some of this written material

I will look at the gathered data to try and develop a global view of students' learning experiences. Any reports of the research findings will thus present broad themes only and the identities of the teacher, students and school are carefully protected. Where selected data from transcribed material is used to support the summary of themes, I will use pseudonyms to again prevent identification. The findings will be presented as part of my doctoral thesis, at seminars and conferences, and published in research journals to help others involved in science education to understand the issues.

If you would like to know more, or meet with me to discuss the project please feel free to contact me. I am most happy to elaborate on any points or discuss any concerns. My contact details are:

Home phone number	07 856 4592	e-mail Hume@hnpl.net
Work phone number	07 8381898	e-mail anne.hume@ero.govt.nz

Regards

Appendix I

Student Questionnaire

Student Questionnaire

During my research I have been trying to discover the things that help and don't help your learning. While I have observed you in class, talked with you and your teacher in interviews and looked at your written work I have noted down the things that appear to affect how you learned. They are listed below. I would like you please to indicate on the scale beside each one how helpful or unhelpful that thing is to your learning.

	1. unhelpful 2. sometimes helpful 3. helpful 4. very helpful
1. Pre-tests	1---4
2. Knowing beforehand what you have to learn	1---4
3. Things you already know	1---4
4. Learning about things that interest you	1---4
5. A knowledgeable teacher	1---4
6. A teacher you like and trust	1---4
7. A teacher who can control the class	1---4
8. Listening to the teacher explain	1---4
9. Understanding what the teacher says	1---4
10. The teacher asking you questions	1---4
11. The teacher answering questions	1---4
12. The teacher giving notes	1---4
13. Lists of definitions	1---4
14. Teacher demonstrations	1---4
15. Tips from the teacher	1---4
16. Diagrams	1---4

17. Brainstorms 1--4

18. Mind maps 1--4

19. Summaries 1--4

20. Cheat sheets 1--4

21. Doing experiments 1--4

22. Doing practice assessments 1--4

23. Being taught the skills needed first 1--4

24. Being taught the background science 1--4

25. The planning template 1--4

26. Feedback from the teacher after marking 1--4

27. Marking other students' assessments 1--4

28. Assessment schedules 1--4

29. Giving feedback to other students 1--4

30. Feedback from other students 1--4

31. Working in groups 1--4

32. Working in pairs 1--4

33. Working alone 1--4

34. The student workbook 1--4

35. Homework 1--4

36. Family 1--4

37. Post tests 1--4

38. Revision 1--4

39. Getting credits for NCEA 1--4

40. Getting merit and excellence
for NCEA 1---4

Is it possible that some of the things above can be both helpful and unhelpful? If so please name them below and we'll discuss them in the interview.

Other comments?

Appendix J

Student Notes for Formative Assessment Task (Case Study A)

Appendix K

Assessment Schedule for Formative Assessment Task (Case Study A)

Appendix L

Student Notes for Summative Assessment Task (Case Study A)

Appendix M

Assessment Schedule for Summative Assessment Task (Case Study A)

Appendix N

Student Instructions for Formative Assessment Task (Case Study B)

Appendix O

Student Journal

Student Journal

(NCEA Science Achievement Standard 1.1. research)

Dear

I am unable to be present in class next week when your teacher hands back your formative assessment and goes over it. I would really appreciate it if you could answer these few questions for me just so I have a record of your experiences and thoughts.

1. How useful was the formative assessment? What have you learned so far?

2. Do you know what you still have to learn to improve your achievement? If so, please describe.

3. What can you do to achieve this extra learning? Please describe.

4. How confident are you feeling about your learning progress? Please explain.

Thank you very much for this helpful information
Anne Hume

Appendix P

Student Instructions for Summative Assessment Task (Case Study B)

Appendix Q

The Modified Science Achievement Standard 1.1 *Carrying out a practical investigation with direction,* as October 2005

Achievement Standard

Subject Reference		Science 1.1			
Title		Carry out a practical science investigation with direction			
Level	1	**Credits** 4		**Assessment**	Internal
Subfield	Science				
Domain	Science – Core				
Registration date	27 October 2004	**Date version published**	27 October 2004		

This achievement standard involves carrying out a practical investigation, with direction, by planning the investigation, collecting and processing the data, and interpreting and reporting the findings.

Achievement Criteria

Achievement	Achievement with Merit	Achievement with Excellence
• Carry out a practical science investigation.	• Carry out a quality practical science investigation.	• Carry out and evaluate a quality practical science investigation.

Explanatory Notes

9 This achievement standard is derived from *Science in the New Zealand Curriculum,* Learning Media, Ministry of Education, 1993, 'Developing Scientific Skills and Attitudes', pp. 42-51; and *Pūtaiao i roto i te Marautanga o Aotearoa*, Learning Media, Ministry of Education, 1996, 'Ngā Pūkenga me Ngā Waiaro ki te Pūtaiao', pp. 70-85.

10 Procedures outlined in *Safety and Science: a Guidance Manual for New Zealand Schools*, Learning Media, Ministry of Education, 2000, should be followed. Investigations should comply with the Animal Welfare Act 1999, as outlined in *Caring for Animals: a Guide for Teachers, Early Childhood Educators, and Students,* Learning Media, Ministry of Education, 1999.

11 An *investigation* is an activity covering the complete process: planning, collecting and processing data, interpreting, and reporting on the investigation. It will involve the student in the collection of primary data.
The investigation will be directed. This means that general instructions for the investigation will be specified in writing and direction will be given in the form of the equipment and/or chemicals from which to choose. A template or suitable format for planning the investigation will be provided for the student to use.

12 Investigations should be based on situations in keeping with content drawn from up to and including science/pūtaiao curriculum Level 6. Possible contexts are given in the curriculum documents.

13 If a student enters for assessment against AS90186, Science 1.1, as well as any of: AS90156, Agriculture and Horticulture 1.1; AS90161, Biology 1.1; AS90169, Chemistry 1.1; or AS90180, Physics 1.1, the investigations must be in different subject areas. For example, if a student is being assessed against AS90161, Biology 1.1, and is also being assessed against AS90186, Science 1.1, then the emphasis of their investigation for AS90186, Science 1.1, cannot be based on biology.

14 A *practical science investigation* will involve:

- a statement of the purpose – this may be an aim, testable question, prediction, or hypothesis based on a scientific idea
- identification of a range for the independent variable or sample
- measurement of the dependent variable or the collection of data
- collecting, recording and processing data relevant to the purpose
- a conclusion based on the interpretation of the processed data.

15 A *quality practical science investigation* enables a valid conclusion to be reached. This would normally involve:

- a statement of the purpose – this may be an aim, testable question, prediction or hypothesis based on a scientific idea
- a method that describes: a valid range for the independent variable or sample; a description of and/or control of other variables; the collection of data with consideration of factors such as sampling, bias, and/or sources of error
- collecting, recording and processing of data to enable a trend or pattern (or absence) to be determined
- a valid conclusion based on interpretation of the processed data that links to the purpose of the investigation.

16 *Evaluate* means to justify the conclusion in terms of the method used. Justification will involve, where relevant, consideration of the:

- reliability of the data
- validity of the method
- science ideas.

Quality Assurance

3 Providers and Industry Training Organisations must be accredited by the Qualifications Authority before they can register credits from assessment against achievement standards.

4 Accredited providers and Industry Training Organisations assessing against achievement standards must engage with the moderation system that applies to those achievement standards.

Accreditation and Moderation Action Plan (AMAP) reference 0226

Appendix R

"How weedy is that field" – Internal Assessment Resource Reference Number: Sci/1/1 BB version 1

National Certificate of Educational Achievement
TAUMATA MĀTAURANGA Ā-MOTU KUA TAEA

2005

Internal Assessment Resource

Subject Reference: **Science 1/1**

Internal assessment resource reference number:
Sci/1/1 – BB version 1

"How weedy is that field"

Supports internal assessment for:

Achievement Standard 90186 version 3
Carry out a practical science investigation with direction.
Credits: 4

Date version published:	January 2005.
Ministry of Education assurance status	For use in internal assessment from 2005.

Teacher Guidelines:

The following guidelines are supplied to enable teachers to carry out valid and consistent assessment using this internal assessment resource. These teacher guidelines do not need to be submitted for moderation.

Context/setting:

This assessment resource is based on planning, carrying out, processing and interpreting, and reporting of a practical investigation that is a **pattern seeking** investigation. The teacher directs what type of investigation the students are to do and changes the planning sheets and student instructions accordingly.

Conditions:

This assessment activity is to be carried out in three parts that lead to the production of an investigation report.
The specific conditions should be stated on the student instruction sheet. e.g. equipment and materials available.

The students need sufficient time for:

- trialling and planning
- carrying out
- processing and interpreting data
- writing a report

The time allowed will depend on the particular investigation chosen. State this time on the student instruction sheet.

Teachers need to be aware of the credit value of this standard when determining the time needed to carry out the investigation.

Any special safety requirements **must** be stated on the student instruction sheet.

Resource requirements:

Students will need to be provided with the materials and equipment required for trialling and carrying out the investigation.

The Investigation

Part 1: Developing a Plan

- The student is provided with a *Planning Sheet* (included) and will work independently to complete this. The planning sheet may need to be modified, related to the task chosen, to allow sufficient space for students to write.
- The student should be given the opportunity to conduct trials to develop their method eg to establish a suitable range of values for the independent variable for a fair test or the sample selection for pattern seeking. A record of this trialling needs to be mentioned on the template or in the final report.
- The student uses the planning sheet and trial results to write a detailed, step-by-step method. The Planning sheet (or other check sheets) may be used to self-evaluate that the method is workable.

Part 2: Collecting and Recording Data

The student follows their written method to collect their own data. The method may be modified but these modifications must be included in their final report and indicated to the assessor.

Part 3: Processing and Interpreting Results

The student must process the data collected into a form that shows a pattern or a trend or absence. This may be achieved by averaging, using a table or using a graph.

Part 4: Presenting a Report

The student, working independently, presents the report of the investigation following the directions/format given in the student instructions.

Teacher Resource Sheet

Prior teaching will need to occur on the scientific method and how to design a practical that involves **pattern seeking.**
Students will need to have been shown how to use quadrats / line transects, how to identify pasture plants and common weeds and how scientists take random samples.

2005

Internal Assessment Resource

Subject Reference: **Science 1.1**

Internal assessment resource reference number: Sci/1/1 – BB version 1

"How weedy is that field"

Achievement Standard 90186 version 3
Carry out a practical science investigation with direction
Credits: 4

Student Instructions

School/Institution	
Student Name	
Teacher or Class reference	
Date of completion	

Background Information:
Your school wants to know if it is time to re-sow with new grass the main playing field. You have found out that when a playing field is about 25% weeds it is time to be re-sown.

In this investigation you are to develop and carry out an investigation. You will plan, collect, process and interpret information, and present a report on how weedy is the main playing field. This investigation is a **pattern seeking** investigation.

Conditions:
This assessment activity is to be carried out in four parts that leads to the production of an investigation report. This investigation must be carried out individually.
Times: This investigation will take 1 week

- trialling and planning 1 period
- carrying out and processing data 1 period
- interpreting the data 1 period
- writing the final report 1 period

All safety procedures **must** be followed, especially when out of the class room.

Equipment:
You have been given some rulers, quadrats, line transects and plant identification charts and books.

Part 1 The Plan

1. State the purpose of your investigation
2. Identify the key variables of the investigation:
   - what will be sampled
   - the sample size
   - other variables that could influence the investigation
3. Describe what will be sampled. Trialling will help you establish this range.
4. Describe how big the sample size will need to be to give valid data.

Controlled variables:

1. Identify any other variables that might influence your investigation and describe how they will be controlled or kept the same to make your results more accurate.
2. Describe how you will ensure that your results are reliable and that you have enough data.

Now write a detailed **step-by-step method** that you will use.

You may change your method as you carry it out as long as you describe any changes made to the method in your report.

Part 2 Collect and Record Data

- Follow your method to collect data and record the results in a table or another appropriate way.
- Remember to record any changes to your method and reasons for the changes as you go.
- Record any difficulties with equipment, gathering your data or your method.

Part 3 Process and Interpret Results

- Process your results so that you can show the trend (or lack of) or pattern in your data. This will usually involve some calculations (e.g. averages) and/or a graph.
- Record the relevant trend or pattern; this is your interpretation.
- Relate the trend or pattern to your purpose; this is your conclusion.

Part 4 Present a Report
Present a report on your investigation. This will include your:

- Trialling and planning sheet
- detailed step-by-step method, including any changes made during your investigation
- recorded data
- processed data
- interpretation of results
- conclusion that links your interpretation to the purpose of the investigation
- evaluation of the conclusion in terms of the method used. In this you may comment on the
- reliability of data (repeats / outliers etc)
- limitations to the investigation (sources of error etc)
- changes made to your original method.
- science ideas related to the investigation. This would explain the results you obtained.

Planning Sheet **Student name**:

1. Purpose of investigation (this may be an aim, testable question, prediction or hypothesis)

2(b). PATTERN SEEKING
What are you going to sample

First variable.

Second variable.

3(b). PATTERN SEEKING
How many samples will you need to take to get reliable data.

Reason:

2. Other variables that need to be controlled to make your results more accurate.

Other Variables	Describe how this variable will be controlled or kept the same?

3. How will you ensure that your results are reliable?

4. Notes from your trials.

Now use the information on this planning sheet to write a detailed step-by-step method.

<u>Method</u>

Changes made to the method after the investigation started.

Report Sheet

Recorded data:

Processed data:

Interpretation of Data:

Conclusion:

Evaluation of the Method and Data and / or Science ideas:

Assessment schedule Sci/1/1 –BB version 1 **How weedy is that field.**

Ēvidence	Achievement	Merit	Excellence
Report contains	Statement of purpose *To find out the percentage of weeds on the playing field.*	(as for achievement)	(as for achievement)
Report (planning sheet and or method)	Identify the 2 key variables. *The weed plants verses the pasture plants.* **And** Measurement of the 2 key variables. *Use quadrats / line transects to take a representative random sample*	achievement plus a workable method (easily followed by another student) is required that includes: • A **valid** range of the sample *25 quadrats / line transects randomly taken across the field* • Description of **and** control of other variables *Eg; non randomness controlled by ensuring following random sampling; paths, leave out of the sampling; poor identification of plants, use good identification techniques.* • Consideration of other factors. *Wrong identification of plants.*	(As for merit)
Ēvidence	Achievement	Merit	Excellence
Report (Recorded data, processed results)	Collect, record and process data relevant to purpose *Some samples taken and weeds V pasture plants identified* *Data recorded in a table or bar graph attempted.* *Trend shown.*	Collect record and process data to enable a **valid** pattern or trend (or absence) *Recorded data processed correctly and bar graph correctly drawn.* *Valid trend correctly links to the data collected.*	(as for merit)

	<u>Possible results.</u> ***Pasture plants Number*** **Grasses 64** **Clovers 28** **Total = 92 (66%)** ***Weeds*** **Plantain 15** **Daisy 14** **Dandelion 12** **Cotula 6** **Total = 47 (34%)**		
Report (Interpretation and conclusion)	A conclusion based on the processed data collected. Eg: the field has about 30% weeds.	A **valid** conclusion that links to the purpose. *Eg:The field has 34% weeds present compared to pasture and therefore should be re-sown*	(as for merit)
Ēvidence	Achievement	Merit	Excellence
Report (evaluation)			Evaluation to justify the conclusion. *A large sample size was taken to make the data reliable.* *Our random number chart put a lot of readings on path areas which were ignored.* *Originally we started using a quadrat but found numbers of plants difficult to identify so changed to a transect.* *Weeds more successful than grasses so take over the grasses with time.* ***Science ideas*** *The weed count in pasture increases due to competition between the weeds and grasses. Once the weed level reaches a certain percentage farmers need to re sow the pasture.*

To determine the overall level of performance all judgements within a column must be met
For each judgement, evidence can be obtained from anywhere in the report

REFERENCES

Adelman, C., Kemmis, S., & Jenkins, D. (1980). Rethinking case study: Notes from the Second Cambridge Conference. In H. Simons (Ed.), *Towards a science of the singular* (pp.45-61). East Anglia: Centre for Applied Research in Education, University of East Anglia.

Alton-Lee, A., & Pratt, A. (2003). *Explaining and Addressing Gender Differences in the New Zealand Compulsory School Sector: A Literature Review.* Retrieved August 8th, 2003 from www.minedu.govt.nz.

Anderson, L.W., & Krathwohl, D. (Eds.). (2001). *A taxonomy for learning, teaching and assessing: A Revision of Bloom's Taxonomy of Educational Objectives.* New York: Longman.

Atkin, J.M., & Black, P. (2003). *Inside science education reform. A history of curricular and policy change.* New York: Teachers College Press.

Austin, L. (2000). What science teachers should know and teach. In G. Haisman (Ed.), *Exploring issues in science education* (pp. 3-10). Wellington, New Zealand: Learning Media.

Baker, R. (1999). Teachers views: Science in the New Zealand Curriculum and related curriculum issues. *New Zealand Science Teacher, 9,* 3-17.

Ball, S. (1984). Beachside reconsidered: Reflections on a methodological apprenticeship. In R. Burgess (Ed.), *The research process in educational settings: Ten case studies* (pp. 69-96). London: Falmer.

Barnett, J., & Hodson, D. (2001). Pedagogical context knowledge: Toward a fuller understanding of what good science teachers know. *Science Education, 85*(4), 426-453.

Beck, R.N. (1979). *Handbook in social philosophy*. New York: Macmillan.

Bell, B., Jones, A., & Carr, M. (1995). The development of the recent national New Zealand science curriculum. *Studies in Science Education, 26*, 73-105.

Bell, B., & Gilbert, J. (1996). *Teacher development: A model from science education*. London: Falmer.

Bell, B., & Baker, R. (Eds.). (1997). *Developing the science curriculum in Aotearoa New Zealand.* Auckland: Longman.

Bell, B., & Cowie, B. (2001). *Formative assessment in science.* Dordrecht: Kluwer.

Bell, B. (2005). Pedagogies developed in the Learning in Science Projects and related theses. *International Journal of Science Education, 27*(2), 159-182.

Bell, J. (1999). *Doing Your Research Project*. Berkshire, United Kingdom: Open University Press.

Biggs, J. (1998). Assessment and classroom learning: A role for summative assessment? *Assessment in Education*, *5*(1), 103-110.

Bishop, R. (1997). Interviewing as collaborative storying. *Education Research and Perspectives, 24*(1), 28-47.

Bishop, R., & Glynn, T. (1999). *Culture counts: Changing power relations in education*. Palmerston North, New Zealand: Dunmore.

Black, P. (1995). Curriculum and assessment in science education: The policy interface. *International Journal of Science Education, 17*(4), 453-469.

Black, P. (2001). Dreams, strategies and systems: Portraits of assessment past, present and future. *Assessment in Education, 8*(1), 65-85.

Black, P. (2003a). *Report to the Q ualifications Development Group Ministry of Education, New Zealand on the proposals for development of the National Certificate of Educational Achievement*. Retrieved March 14, 2003 from http://www.minedu.govt.nz.

Black, P. (2003b, April*). Formative and summative assessment: Can they serve learning together?* Paper presented at American Education Review Association, SIG Classroom Assessment Meeting 52.028.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, *5*(1), 7-74.

Bloom, B., Englehart, M., Furst, E., Hill, W., & Krathwohl, D. (1956). Taxonomy of educational objectives: The classification of educational goals. *Handbook I: Cognitive domain*. New York: Longmans Green.

Blundell, S. (2003, February 15-21). Can do better. *The Listener*, 30-31.

Brookhart, S. M. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education, 8*(2), 153-170.

Brookhart, S. M., & Bronowicz, D.L. (2003). 'I don't like writing. It makes my fingers hurt': Students talk about their classroom assessments. *Assessment in Education*, *10*(2), 221-242.

Brown, J., Collins, A., & Duguid, P. (1980). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42.

Bullock, A., & Trombley, S. (2000). *Fontana dictionary of modern thought* (3rd ed.). Glasgow: Harper-Collins.

Burns, R.B. (1994). *Introduction to research methods* (2nd ed.). Melbourne, Australia: Longman.

Burt, E., & Davison, L. (1998). Teacher professional development: Then, now and in the future. *New Zealand Journal of Educational Administration, 13,* 44-52.

Butler, D.L. & Winne, P.H. (1995). Feedback and self-regulated learning: a theoretical synthesis. *Review of Educational Research*, 65(3), pp. 245-281.

Cameron, M. (2002). Peer influences on learning. *Set: Research Information for Teachers 3*, 36-40.

Carlson Powell, J., & Anderson, R. D. (2002). Changing teachers' practice: Curriculum materials and science education reform in the USA. *Studies in Science Education, 37,* 107-136.

Carr, M. (2001). A sociocultural approach to learning orientations in an early childhood setting. *Qualitative Studies in Education, 14,* 525-542.

Carr, M., McGee, C., Jones, A., McKinley, E., Bell, B., Barr, H., & Simpson, T. (2001). *The effects of curricula and assessment on pedagogical approaches and on educational outcomes.* Retrieved March 14, 2003, from http://www.minedu.govt.nz

Chin, C., & Kayalvizhi, G. (2002). Posing problems for open investigations: What questions do students ask? *Research in Science & Technological Education, 20*(2), 269-287.

Clarke, S. (2001). *Unlocking formative assessment. Practical strategies for enhancing pupils' learning in the primary classroom*. London: Hodder & Stoughton.

Cohen, L., Manion, L. & Morrison, K. (2000). *Research methods in education* (5th ed.). London: Routledge.

Cole, M., & Engestrom, Y. (1993). A cultural-historical approach to distributed cognition. In G. Salomon (Ed.). *Distributed cognitions: Psychological and educational considerations*. Cambridge, UK: Cambridge University.

Coles, B., & Coles, C. (1997). Investigating with CREST. In B. Bell & R. Baker (Eds.), *Developing the science curriculum in Aotearoa New Zealand* (pp. 81-90). Auckland: Longman.

Coll, R. (2000). The importance of triangulation for research in science education. In R.K. Coll (Ed.), *SAMEpapers* (pp. 31-40). Hamilton, New Zealand: Centre for Science and Technology Education Research, University of Waikato.

Coll, R. (2002). *Multiple approaches for research in education*. Hamilton, New Zealand: School of Education, University of Waikato.

Collins, N. (2004). Scientific research and school students. *School Science Review*, *85*(312), 77-86.

Cooper, B., Hume, A., & Abbott, G. (2002). *Year 11 science. NCEA level 1 workbook.* Hamilton, New Zealand: ABA.

Cowie, B., & Bell, B. (1999). A methodology for investigating formative assessment in the science classroom. In F. Biddulph & K. Carr (Eds.), *SAMEpapers (*pp. 36-62). Hamilton, New Zealand: Centre for Science, Mathematics and Technology Education Research, University of Waikato.

Cowie, B., Jones, A., & Harlow, A., (2003). Primary and secondary teachers experiences of Science in the New Zealand Curriculum. *New Zealand Science Teacher, 104*, 29-34.

Crooks, T. J. (1998). The impact of classroom evaluation practices on students. *Review of Educational Research*, *58*(14), 438-481.

Crooks, T. (2002a, December). *Assessment, accountability and achievement – principles, possibilities and pitfalls.* Keynote address 24th annual conference of the New Zealand Association for Research in Education, Palmerston North.

Crooks, T. (2002b). Educational assessment in New Zealand schools. *Assessment in Education*, *9*(2), 237-253.

Crooks, T., & Flockton, L. (2000a). Graphs, tables and maps: Assessment results 1999. *National Education Monitoring Report 15*. Dunedin, New Zealand: The Educational Assessment Research Unit, University of Otago.

Crooks, T., & Flockton, L. (2000b). Science assessment results 1999. *National Education Monitoring Report 13*. Dunedin, New Zealand: The Educational Assessment Research Unit, University of Otago.

Crooks, T., & Flockton, L. (2004). Science assessment results 2003. *National Education Monitoring Report 29*. Dunedin, New Zealand: The Educational Assessment Research Unit, University of Otago.

Davies, J., & France, B. (2001). *Science and technology extension in schools: CREST and the Olympiads.* Final report to the Ministry of Research, Science and Technology, New Zealand.

Deboer, G. (2002). Student-centred teaching in a standards-based world: Finding a sensible balance. *Science & Education, 11*, 405-417.

Dewar, S., & Bennie, N. (1996). An evaluation of some of the professional programmes delivered in 1995/6. *The Research Bulletin, 7,* 68-77.

Dixon, H., & Williams, R. (2003). Teachers' understanding and use of formative assessment in literacy learning. In I. Livingstone (Ed.), *New Zealand Annual Review of Education* (pp. 95-110). Wellington, New Zealand: School of Education, Victoria University of Wellington.

Driver, R., Leach, J., Millar, R., & Scott, P. (1996). *Young people's images of science*. Buckingham: Open University Press.

Duggan, S., & Gott, R. (1995). The place of investigations in practical work in the UK National Curriculum for Science. *International Journal of Science Education, 17*(2), 137-147.

Duggan, S., & Gott, R. (2002). What sort of science education do we really need? *International Journal of Science Education*, *24*(7), 661-680.

Duschl, R.A., & Hamilton, R, J. (1998). Conceptual change in science and in the learning of science. In B.J. Fraser & K.G.Tobin (Eds.), *International handbook of science education* (pp. 1047-1065). Dordrecht: Kluwer.

Education Review Office. (1996). *Science in schools: implementing the 1995 curriculum.* Wellington, New Zealand: Education Evaluation Reports.

Education Review Office. (2000). In-service training for teachers in New Zealand schools. *Education Evaluation Reports*. Wellington, New Zealand: Crown Copyright.

Education Review Office. (2002). *Ethics and conflicts of interest.* Wellington, New Zealand: Crown Copyright.

Education Review Office. (2002). *Science now: science education for students in Years 1 to 8.* Wellington, New Zealand: Education Evaluation Reports.

Education Review Office. (2003). *Evaluation indicators for education reviews in schools.* Wellington, New Zealand: Crown Copyright.

Education Review Office. (2004). *Education Review Office evaluation report. Progress of NCEA levels 1and 2, and readiness of schools to implement level 3.* Wellington, New Zealand: Crown Copyright.

Eisenhart, M.A., & Howe, K.R. (1992). Validity in educational research. In M.D. LeCompte, W. Millroy, & J. Priessle, (Eds.), *The handbook of qualitative research in education* (pp. 643-680). San Diego, CA: Academic Press.

Eisner, E. (1983). Anastasia might still be alive, but the monarchy is dead. *Educational Researcher*, *12*(5), 13-34.

Eisner, E. (1994). *The educational imagination* (3^{rd} ed.). New York: Macmillan.

Elley, W. (2003, February 19-25). New assessment system does not pass test. *New Zealand Education Review*, 5-6.

English, E. (1997). The interpretation of the curriculum. Is it all a matter of serendipity? *New Zealand Science Teacher, 85,* 27-33.

English, C., & Wood, R. (1997). The doing of science – a redefinition of procedural knowledge. In B. Bell & R. Baker (Eds.), *Developing the science curriculum in Aotearoa New Zealand* (pp. 67-80). Auckland: Longman.

Engstrom, Y., Engstrom, R., & Sunito, A. (2002). Can a school community learn to master its own future? An activity-theoretical study of expansive learning among middle school teachers. In G. Wells & G. Claxton (Eds.), *Learning for life in the 21st century. Sociocultural perspectives on the future of education* (pp. 211-224). Oxford: Blackwell.

Erickson, F., & Shultz, J. (1992). Students' experience of the curriculum. In P.W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 465-485). New York: Macmillan.

Erikson, F. (1998). Qualitative research methods for science education. In B.J. Fraser & K.G. Tobin (Eds.), *International handbook of science education* (pp. 1155-1174). Dordrecht: Kluwer.

Fancy, H. (2001, November). Howard Fancy thinks…. S*panzjournal,* November, 3-6.

Forret, J. (1999). Interdisciplinary research: Science and law. In F. Biddulph & K. Carr (Eds.), *SAMEpapers* (pp. 63-77). Hamilton, New Zealand: Centre for Science, Mathematics and Technology Education Research, University of Waikato.

Foster, G. (2000). At last! The chance to assess the 'doing' of science. *New Zealand Science Teacher, 93,* 33-38.

France, B. (2001). Biotechnology teaching models. What is their role in technology education? In R.K. Coll (Ed.), *SAMEpapers* (pp. 34-52). Hamilton, New Zealand: Centre for Science and Technology Education Research, University of Waikato.

Garnett, P.J., & Garnett, P. J. (1995). Refocussing the chemistry lab: A case for laboratory-based investigations. *Australian Science Teachers' Journal, 41*(2), 26-34.

Gilmore, A. (1994). *Mathematics and science in the national curriculum: evaluation of the teacher development programmes*. Christchurch, New Zealand: Education Department, University of Canterbury.

Gipps, C. (1994). *Beyond testing: towards a theory of educational assessment.* London: Falmer.

Glaser, B.G., & Strauss, A.L. (1967). *The discovery of grounded theory*. Chicago: Aldine.

Glover, P., & Thomas, R. (1999). Coming to grips with continuous assessment. *Assessment in Education, 6*(1), 117-127.

Goetz, J., Preissle., & LeCompte, M.D. (1984). *Ethnography and qualitative design in educational research*. Orlando, FL: Academic Press.

Gott, R., & Duggan, S. (2002). Problems with the assessment of performance in practical science: Which way now? *Cambridge Journal of Education, 32*(2), 183-201.

Gott, R.; Duggan, S., & Johnson, P. (1999). What do practising applied scientists do and what are the implications for science education? *Research in Science and Technological Education, 17*(1), 97-107.

Graue, E., & Walsh, D. (1999*). Studying children in context. Theories, methods and ethics.* London: Sage.

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

Haigh, M. (1995). The Achievement Initiative in Science: Developing Science in the New Zealand Curriculum. In A. Jones, A. Begg, B. Bell, F. Biddulph, M.Carr, M. Carr, J. McChesney, E. McKinley & J. Young Loveridge (Eds.), *SAMEpapers* (pp. 45-74). Hamilton, New Zealand: Centre for Science, Mathematics and Technology Education Research, University of Waikato.

Haigh, M., & Hubbard, D. (1997). ‘I really know I have learned something’: investigative work in science education. In B. Bell & R. Baker (Eds.). *Developing the science curriculum in Aotearoa New Zealand* (pp. 53-66). Auckland: Longman.

Haigh, M. (2000). The many faces of case study research. In R.K. Coll (Ed.), *SAMEpapers* (pp. 69-93). Hamilton, New.Zealand: Centre for Science and Technology Education Research, University of Waikato.

Haigh, M. (2001). Supporting student investigations. *New Zealand Science Teacher, 96*, 39-41.

Haig, M. (2003). “Miss, miss, I’m thinking, I’m thinking”: fostering creativity through science education. *New Zealand Science Teacher, 105,* 8-10.

Haigh, M., France , B., & Forret, M. (2005). Is ‘doing science’ in New Zealand classrooms an expression of scientific inquiry? *International Journal of Science Education*, *27*(2), 215-226.

Hannay, B., Howison, P., & Sayes, M. (2002). *Year 11 Science. Study guide. NCEA Level 1 edition.* Auckland: ESA

Harding, S. (Ed.). (1987). *Feminism and methodology*. Bloomington, IN: Indiana University Press.

Hargreaves, A., & Fullan, M.G. (Eds.). (1992). *Understanding teacher development*. London: Cassell.

Harlen, W., & James, M. (1997). Assessment and learning: differences and relationships between formative and summative assessment. *Assessment in Education, 4*(3), 365-379.

Harlen, W. (1999). Purposes and procedures for assessing science process skills. *Assessment in Education*, *6*(1), 129-144.

Harlen, W. (1998). The last 10 years: the next 10 years. In R. Sherrington (Ed.). *ASE guide to primary science education.* (pp. 23-33). Cheltenham, UK: Stanley Thornes for the Association for Science Education.

Harlen, W., & Crick, R.D. (2003). Testing and motivation for learning. *Assessment in Education*, *10*(2), 169-207.

Hart, C., Mulhall, P., Berry, A., Loughran, J., & Gunstone, R. (2000). What is the purpose of this experiment? Or can students learn something from doing experiments? *Journal of Research in Science Teaching, 37*(7), 655-675.

Hattie, J. (2002). Assessment in New Zealand. *Assessment in Education, 9*(2), 237-253.

Hennessey, S. (1993). Situated cognition and cognitive apprenticeship: implications for classroom learning. *Studies in Science Education, 22,* 1-41.

Hipkins, R., Bolstad, R., Baker, R., Jones, A., Barker, M.; Bell, B., Coll, R.K., Cooper, B., Forret, M., Harlow,A., Taylor, I., France, B., & Haigh, M. (2001). *Curriculum, learning and effective pedagogy: a literature review in science education.* Wellington, New Zealand: Ministry of Education.

Hipkins, R., & Barker, M. (2002). Science in the New Zealand curriculum: present potential and future possibilities. *New Zealand Science Teacher, 100*, 10-16.

Hipkins, R., & Booker, F. (2002). You can't investigate in a vacuum. S*et: Research Information for Teachers 3,* 22-26.

Hipkins, R., & Vaughan, K. (2002). *Learning curves: meeting student needs in an evolving qualifications regime. From cabbages to kings: a first report*. Wellington: New Zealand Council for Educational Research.

Hipkins, R.,Vaughan, K., Beals, F., & Ferral H.(2004). *Learning curves: meeting student learning needs in an evolving qualifications regime. Shared pathways and multiple tracks: A second report*. Wellington: New Zealand Council for Educational Research.

Hipkins, R. (2004, July). *Shifting balances : A study of the impact of assessment for qualifications on classroom teaching.* Paper presented at the annual conference of the Australian Science Education Research Association, Armidale.

Hipkins, R., & Vaughan, K. (2005). *Shaping our futures: meeting secondary students' learning needs in a time of evolving qualifications. The third and final report.* Wellington: New Zealand Council for Educational Research.

Hodson, D. (1990). Practical work in school science. *School Science Review*, *71*(256), 33-40.

Hodson, D. (1991). Practical work in science: time for a reappraisal. *Studies in Science Education*, *19,* 175-184.

Hodson, D. (1992). Assessment of practical work: some considerations in philosophy of science. *Science & Education, 1*, 115-144.

Hodson, D. (1996). Laboratory work as scientific method. *Journal of Curriculum Studies, 28*(2), 115-135.

Hughes, P. (2004). Mainstream chemical research in school science. *School Science, Review*, *85*(312), 71-76.

Hume, A.C. (1991). Engaging students in science: a case of negotiated curriculum development from earth sciences. In A. Begg, B. Bell, F. Biddulph, M. Carr, M Carr, J. McChesney, & J. Young-Loveridge, *SAMEpapers* (pp. 205-224). Hamilton, New Zealand: Centre for Science, Mathematics and Technology Education Research, University of Waikato.

Hume, A.C. (1995). Incorporating CREST into your chemistry course. *New Zealand Science Teacher*, *79,* 32-36.

Hurd, P. (1997). Scientific literacy: new minds for a changing world. *Science Education*, *82*(3), 407-416.

Jenkins, W.J. (1996). The ‘nature of science’ as a curriculum component. *Journal of Curriculum Studies*, *28*(2), 137-150.

Kang, N., & Wallace, C.S. (2004). Secondary science teachers’ use of laboratory activities: linking epistemological beliefs, goals, and practices. *Science Education*, *89*(1), 140-165.

Keeves, J.P. (1998). Methods and processes in research in science education. In B. Fraser & K. Tobin. (Eds.). *International handbook of science education* (pp. 1127-1153). Dordrecht: Kluwer.

Keiler, L.S., & Woolnough, B.E. (2002). Practical work in school science: the dominance of assessment. *School Science Review*, *83*(304), 83-88.

Kemmis, S., & McTaggart, R. (Eds.). (1988). *The action research planner* (3rd ed.). Victoria, Australia: Deakin University Press.

Kennedy, M. (1997). The connection between research and practice. *Educational Researcher, 26*(7), 4-12.

Kerlinger, F.N. (1986). *Foundations of behavioural research* (3rd ed.). New York: Rinehart & Winston.

Knapp, M.S. (2002). *Understanding how policy meets practice: two takes on local response to a state reform initiative. An occasional paper.* Washington, DC: Center for the Study of Teaching and Policy, University of Washington.

Kuhn, T.S. (1962). *The structure of scientific revolutions*. Chicago & London: The University of Chicago Press.

Lake, D. (2004). What makes a student's science investigation more 'scientific'. *School Science Review, 85*(312), 107-111.

Lather, P. (1992). Critical frames in educational research: feminist and post-structural perspectives. *Theories into Practice, 13*(2), 87-99.

Laws, P.M. (1996). Investigative work in the Science National Curriculum. *School Science Review*, *77*(281), 17-25.

Leach, J., & Scott, P. (2003). Individual and sociocultural views of learning in science education. *Science and Education, 12*, 91-113.

Lederman, N.G. (1999). Teachers' understanding of the nature of science and classroom practice: factors that facilitate or impede the relationship. *Journal of Research in Science Teaching, 36*(8), 916-929.

Lee, H., & Lee, G. (2001). The National Certificate of Educational Achievement (NCEA): 'Fragile - handle with care'. In I. Livingstone (Ed.), *New Zealand Annual Review of Education* (pp. 5-37). Wellington: School of Education, Victoria University of Wellington.

Lennox, B. (1995). *School assessment for the National Qualifications Framework.* Unpublished research paper for Master of Education, Victoria University, Wellington.

Lewthwaite, B., (1999). Implementing the New Zealand science curriculum: will there be a change of state? *New Zealand Science Teacher, 92*, 6-10.

Lewthwaite, B. (2000). Implementing science in the New Zealand curriculum: how teachers see the problems. In *Exploring Issues in Science Education* (pp. 11-22). Wellington, New Zealand: Research and Curriculum Divisions, Ministry of Education.

Lewthwaite, B., Stableford, J., & Fisher, D. (2001). Enlarging the focus on primary science delivery New Zealand. In R.K. Coll (Ed.), *SAMEpapers* (pp. 213-237). Hamilton, New Zealand: Centre for Science and Technology Education Research, University of Waikato.

Linsell, C. (1999). Developing a methodology for investigating learning about probability in a constructivist classroom. In F. Biddulph & K. Carr (Eds.), *SAMEpapers* (pp. 118-137). Hamilton, New Zealand: Centre for Science and Technology Education Research, University of Waikato.

Lokan, J., Adams, R., & Doig, B. (1999). Broadening assessment, improving assessment? Some examples from school science. *Assessment in Education, 6*(1), 83-99.

Loughran, J., & Gunstone, R. (1997). Professional development in residence: developing reflection on science teaching and learning. *Journal of Education for Teaching, 23*(2), 159-178.

Loveless, M., & Barker, M. (2000). Those pages we just turn over….: The nature of science in Science in the New Zealand Curriculum. *New Zealand Science Teacher, 93*, 28-31.

Luft, J. (1999). Challenging myths. Questioning common myths about science education. *The Science Teacher, 66*(4), 40-43.

Mathews, M. (1995). *Challenging New Zealand science education*. Palmerston North, New Zealand: Dunmore.

Mathews, P.S.C., & McKenna, P.J. (2005). Assessment of practical work in Ireland. *International Journal of Science Education, 27*(10), 1211-1224.

Mayer, V., & Kumano, Y. (1999). The role of system science in future school science curricula. *Studies in Science Education, 34,* 71-91.

McBride, T. (1994). The information privacy principles. In E.Longworth & T. McBride (Eds.), *The Privacy Act: A Guide,* (pp. 49-70). Wellington: GP Publications.

McGee, C. (1997). *Teachers and curriculum decision-making*. Palmerston North, New Zealand: Dunmore.

McGee, C., & Penlington, C. (2000). *Research on learning, curriculum and teachers' roles. Report 1: historical overview.* Hamilton New Zealand: Waikato Institute for Research in Learning and Curriculum, University of Waikato.

McGee, C., & Penlington, C. (2001a). *Research on learning, curriculum and teachers' roles. Report 2: teacher-student interaction.* Hamilton New Zealand: Waikato Institute for Research in Learning and Curriculum, University of Waikato

McGee, C., & Penlington, C. (2001b). *Research on learning, curriculum and teachers' roles. Report 3: the classroom curriculum*. Hamilton New Zealand: Waikato Institute for Research in Learning and Curriculum, University of Waikato.

McGee, C., & Penlington, C. (2001c). *Research on learning, curriculum and teachers' roles. Report 4: teacher knowledge, belief and thinking.* Hamilton

New Zealand: Waikato Institute for Research in Learning and Curriculum, University of Waikato

McMillan, B. (1995). Making sense of the debate about the science curriculum. In A. Jones, A. Begg, B. Bell, F. Biddulph, M.Carr, M. Carr, J. McChesney, E. McKinley & J. Young-Loveridge (Eds.), *SAMEpapers* (pp 95-111). Hamilton, New Zealand: Centre for Science, Mathematics and Technology Education Research, University of Waikato

Ministry of Education. (1993a). *Science in the New Zealand curriculum*. Wellington, New Zealand: Learning Media.

Ministry of Education. (1993b). *The New Zealand curriculum framework*. Wellington, New Zealand: Learning Media.

Ministry of Education. (1994a). *Assessment. Policy to practice.* Wellington: Learning Media.

Ministry of Education. (1994b). *Te Tauaki Marautanga Putaiao: he tauira*. Wellington: Learning Media.

Ministry of Education. (1997). *A future qualifications policy for New Zealand: a plan for the National Qualifications Framework*. Wellington: Ministry of Education.

Ministry of Education. (2002a). *Curriculum stocktake report to Minister of Education*. Retrieved September 7, 2003 from http://www.minedu.govt.nz

Ministry of Education. (2002b). *Programme in international student assessment – PISA 2000.* Retrieved August 8, 2003 from http://www.mindeu.govt.nz

Ministry of Education. (2002c). *The school and classroom context for Year 9 students' mathematics and science achievement: results from New Zealand's participation in the repeat of the Third International Mathematics and Science study*. Retrieved August 8, 2003 from http://www.mindeu.govt.nz

Ministry of Education. (2003a). *New Zealand Curriculum Project.* Retrieved September 7, 2003 from http://www.tki.org.nz

Ministry of Education. (2003b). *Teachers' experiences in curriculum implementation: English, languages, science and social studies*. Retrieved 9 October, 2003 from http://www.mindeu.govt.nz

Ministry of Education. (2003c). *Effective literacy practice in Years 1 to 4*. Wellington, New Zealand: Learning Media.

Ministry of Education. (2004a). *Curriculum/marautanga project - focusing on what matters.* Issue 1, April.

Ministry of Education. (2004b). *Making a bigger difference for all students. A schooling strategy discussion document.* Wellington, New Zealand: Medium Term Strategy, Ministry of Education.

Ministry of Education. (2005a). *Science. Draft for Internal MOE Audit* – 24 August 2005. Retrieved 15 November 2005 from www.tki.org.nz.

Ministry of Education. (2005b). *Setting the direction for learning. The New Zealand Curriculum Marautanga Project.* Retrieved 16 October, 2005 from www.tki.org.nz.

Moreland, J., Jones, A., & Chambers, M. (2001). Enhancing student learning in technology through enhancing teacher formative interactions. *Set: Research Information for Teachers, 3*,16-19.

Murphy, P.F. (2003, October). *Changing curriculum, changing teaching - messages emerging from evaluations of curriculum innovations in design and technology.* Keynote paper presented at the New Zealand International Conference for Technology Education, Hamilton, New Zealand.

New Zealand Association of Science Educators. (2003). *SCICON 2002 conference programme*. Retrieved 27 October 2003 from http://www.nzase.org.nz

New Zealand Qualifications Authority. (1996). *Learning and assessment. A guide to assessment for the National Qualifications Framework*. Wellington: New Zealand Qualifications Authority.

New Zealand Qualifications Authority. (2000). *The National Certificate of Educational Achievement professional development programme. Regional facilitator's guide.* Wellington: New Zealand Qualifications Authority.

New Zealand Qualifications Authority. (2001*). The National Certificate of Educational Achievement. An introduction for parents and students.* Wellington: New Zealand Qualifications Authority.

Neyland, J. (1995). Neo-behaviourism and social construction in mathematics education. In A. Jones, A. Begg, B. Bell, F. Biddulph, M.Carr, M. Carr, J. McChesney, E. McKinley & J. Young-Loveridge (Eds.), *SAMEpapers* (pp. 114-143). Hamilton, New Zealand: Centre for Science, Mathematics and Technology Education Research, University of Waikato.

Nikander, P. (1995). The turn to the text: the critical potential of discursive social psychology. *Nordiske Udkast, 2*, 3-15.

Nitko, A.J. (1995). Curriculum-based continuous assessment: a framework for concepts, procedures and policy. *Assessment in Education, 2*(3) 321-338.

Nuthall, G. (1997). Understanding student thinking and learning in the classroom. In B. Biddle, T. Good, & I. Goodson (Eds.), *The International Handbook of Teachers and Teaching* (pp. 1-90). Dortrecht: Kluwer.

Orpwood, G. (2001). The role of assessment in science curriculum reform. *Assessment in Education*, *8*(2), 135-151.

Osborne, J. (2002). Science without literacy: a ship without a sail? *Cambridge Journal of Education*, 23(2), 203-218.

Osborne, J. (2004, January, 2). Clumsy tests dull passion for science. *Times Educational Supplement*, p. 1 .

Patton, M.Q. (1990). *Qualitative evaluation and research methods*. Newbury Park, CA: Sage.

Peddie, R. (1992). *Beyond the norm? An introduction to standards-based assessment.* Wellington: New Zealand Qualifications Authority.

Peterson, P. (1998). Why do educational research? Rethinking our roles and identities, our texts and contexts. *Educational Researcher, 27*(3), 4-10.

Pitman, M.A., & Maxwell, J.A. (1992). Qualitative approaches to evaluation. In M.D. LeCompte, W. Millroy, & J. Priessle. (Eds.). 1992. *The Handbook of Qualitative Research in Education.* (pp. 729-770). San Diego, CA: Academic Press.

Poskitt, J. (1992). *Report on the evaluation of the ACCO teacher development contracts.* Palmerston North, New Zealand: Massey University.

Powell, C.P., & Anderson, R.D. (2002). Changing teachers' practice: curriculum materialsand science education reform in the USA. *Studies in Science Education*, 37, 107-136.

Potter, J. (1998). Qualitative and discourse analysis. In N. Schooler (Ed.), *Comprehensive clinical psychology* (pp. 117-144). Amsterdam: Elsevier Science.

Preece, F.W., & Skinner, N.C. (1999). The National Assessment in Science at Key Stage 3 in England and Wales and its impact on teaching and learning. *Assessment in Education*, *6*(1), 11-25.

Reid, N., & Yang, M-J. (2002). Open-ended problem solving in school chemistry: a preliminary investigation. *International Journal of Science Education*, *24*(12), 1313-1332.

Reay, D., & Wiliam, D. (1999). 'I'll be a nothing': Structure, agency and the construction of identity. *British Educational Research Journal*, *25*(3), 343-354.

Rennie, L.J., Goodrum, M., & Hackling, M. (2001). Science teaching and learning in Australian schools: results of a national study. *Research in Science Education*, *31*(4), 455-498.

Roberts, R., & Gott, R. (2004). Using different types of practical within a problem-solving model of science. *School Science Review*, *85*(312), 113-119.

Rudolph, J. (2002). Portraying epistemology: school science in historical context. *Science Education*, *87*(1), 64-79.

Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18,* 119-144.

Sadler, R. (1998). Formative assessment: revisiting the territory. *Assessment in Education*, *5*(1), 77-84.

Sandoval, W.A. (2005). Understanding students' practical epistemologies and their influence on learning through inquiry. *Science Education*, *89*(4), 634-656.

Schubert, W.H. (1986). *Curriculum: perspective, paradigm and possibility.* New York: Macmillan.

Scott, C. (2000). Qualifications changes … again! *New Zealand Science Teacher, 96,*15-16.

Shulman, L.S. (1981). Disciplines of inquiry in education: an overview. *Educational Researcher*, *10*(6), 5-12.

Skamp, K. (2004). *Teaching Primary Science Constructively* (2nd ed.). Victoria, Australia: Thomson.

Snook, I. (1998, December). *The ethics and politics of research*. Paper presented at the annual New Zealand Association of Researchers in Education conference. Dunedin.

Solano-Flores, G., Jovanovic, J., Shalvenson, R.J., & Bachman, M. (1999). On the development and evaluation of a shell for generating science performance assessments. *International Journal of Science Education*, *21*(3), 293-315.

Spillane, J.P. (2004). *Standards deviation: how schools misunderstand education policy*. Cambridge, MA: Harvard University Press.

Spindler, G., & Spindler, L. (1992). Cultural process and ethnography: An anthropological perspective. In M.D. LeCompte, W. Millroy, & J. Priessle. (Eds.). 1992. *The handbook of qualitative research in education* (pp. 643-680). San Diego, California: Academic Press.

Strachan, J. (2001). Assessment in change: some reflections on the local and international background to the National Certificate of Educational Achievement (NCEA). *New Zealand Annual Review of Education, 11,* 245-262.

Thornton, K. (2003). Teacher refresher courses and effective professional development for teachers. In I. Livingstone (Ed.), *New Zealand Annual Review of Education* (pp. 75-94). Wellington, New Zealand: School of Education, Victoria University.

Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: using action research to explore and modify theory. *British Educational Research Journal*, *27*(5), 615-631.

Tsai, C. (2002). Nested epistemologies: science teachers' beliefs of teaching, learning and science. *International Journal of Science Education*, *24*(8), 771-783.

Tytler, R. (2003). A window for a purpose: developing a framework for describing effective science teaching and learning. *Research in Science Education, 33*, 273-298.

Tytler, R.; & Swatton, P. (1992). A critique of Attainment Target 1 based on case studies of students' investigations. *School Science Review, 74*(266), 21-35.

University of Waikato. (2001). *Handbook on ethical conduct in research*. Hamilton, New Zealand: Vice-chancellors Office.

Wallace, J. & Louden, W. (Eds.). (2002). *Dilemmas of science teaching. Perspectives on problems of practice.* London: Routledge Falmer.

Walshaw, M. (2001). Engaging postpositivist perspectives in mathematics education. In R.K. Coll (Ed.), *SAMEpapers* (pp. 19-33). Hamilton, New Zealand: Centre for Science, Mathematics and Technology Education Research, University of Waikato.

Watson, R., Goldsworthy, A.. Wood-Robinson, V. (1999). What is not fair with investigations? *School Science Review, 80*(292), 101-106.

Weinburgh, M. (2003). Confronting and changing middle school teachers' perceptions of scientific methodology. *School Science and Mathematics, 103*(5), 222-232.

Welch, D. (2003). Done too much, much too young. *The Listener*, pp. 16-21.

Welch, D. (2004). Lesson at school. *The Listener*, pp. 22-24.

Wenger, E. (1998*). Communities of practice. Learning, meaning and identity*. Cambridge: University Press.

White, B.Y., & Fredericksen, J.R. (1998). Inquiry, modeling and metacognition: Making science accessible to all students. *Cognition and Instruction*, *16*(1), 3-118.

Wiliam, D. (2000, August). *Integrating formative and summative functions of assessment.* Paper presented to Working Group 10 of the International Congress on Mathematics Education, Makuhari, Tokyo.

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education, 11*(1), 49-65.

Woolnough, B. (2001). What really happens when students 'do science'- and how we can find out? *Studies in Science Education, 36*, 162-168.

Woolnough, B., & Allsop, T. (1985). *Practical work in science.* Cambridge, UK: Cambridge University Press

Yin, R.K. (2003). *Applications of case study research* (2nd ed.). Thousand Oaks, CA: Sage.

Zevenbergen, R., & Begg, A. (1999). Theoretical frameworks in educational research. In F. Biddulph & K. Carr (Eds.), *SAMEpapers* (pp. 170-185). Hamilton, New Zealand: Centre for Science, Mathematics and Technology Education Research, University of Waikato.