Assessment and Moderation of the Level 2 Physics Unit Standards on the National Qualifications Framework

A thesis

submitted in partial fulfilment

of the requirements for the Degree

of

Doctor of Philosophy

in the

University of Canterbury

by

John Boereboom

University of Canterbury

1999

Assessment and Moderation of the Level 2 Physics Unit Standards on the National Qualifications Framework

John Boereboom

Abstract

The introduction of the National Qualifications Framework and the associated assessment against Physics Unit Standards represents a major paradigm shift in senior secondary school assessment. The trend away from norm-referenced external national exams towards internal standards-based assessment has significant implications for curriculum delivery, student learning and assessment and moderation practices.

The New Zealand Qualifications Authority claims that the National Qualifications Framework is a technically sound and publicly acceptable alternative to the established system. Moderation is a key plank of the Framework which aims to establish and maintain national consistency of assessment across different providers, improve assessment practices, assist in the development of assessor expertise and establish public confidence in the new qualifications.

Critics have expressed doubts that the perceived national consistency and public confidence in national examinations may be lacking in internal assessment against Unit Standards. The Post Primary Teachers' Association has expressed concern that the workload associated with the assessment, reassessment and administration of Unit Standards will adversely affect the quality of curriculum delivery and student learning. The present research is an attempt to evaluate these claims and counterclaims.

In this thesis the assessment and moderation of the Physics Unit Standards is used as a context to answer the question:

> Is assessment against the Physics Unit Standards a valid, reliable and manageable way of assessing the achievement objectives of *Physics in the New Zealand curriculum*?

A range of qualitative and quantitative techniques was employed to monitor the quality assurance of assessment and moderation of the physics Unit Standards over a three-year period and describe its impact on teachers and students.

The research established that assessment against the Physics Unit Standards was generally valid but that doubts remain about its suitability to assess conceptual learning, the micro-definition of learning outcomes and the lack of recognition of different levels of achievement. The moderation action plan was found to be effective in establishing and maintaining satisfactory comparability between schools. After the initial implementation period the workload was manageable for teachers but concerns remained about dual assessment and excessive administrative requirements. Teachers felt that the resources provided by the NZQA were generally adequate.

The research identified a number of areas for improving the quality of assessment against the Level 2 Physics Unit Standards. Chief recommendations related to the incorporation of higher level skills, a broadening of the performance criteria, the recognition of levels of achievement including excellence, the elimination of dual assessment and a reduction in the amount of assessment.

Acknowledgements

I am grateful to the Christchurch College of Education for the payment of university fees and a research grant which helped defray the expenses associated with this kind of research. I should like to acknowledge the New Zealand Qualifications Authority for providing access to their national database and moderation system and their interest in and support of this project.

My supervisors, Dr Alison Gilmore and David Hughes of the Education Department, Canterbury University helped me to transform an initially vague concept into a detailed set of research questions and provided excellent support throughout the period of the research. Dr Cathy Wylie from the New Zealand Council of Educational Research provided helpful feedback in the early stages of the research.

I should also like to thank my wife and children without whose support and tolerance I would have been unable to complete this project.

A final word of thanks to all the physics teachers, students and moderators, who willingly gave of their time over a three year period at often busy times in the school year.

Table of contents

Chapter 4 Validity of Year 12 Physics Assessment

Chapter 5 Moderation of the physics Unit Standards (1996-1998)

Chapter 6 Manageability of Year 12 Physics Assessment

List of Tables

List of Figures

List of Appendices

Abbreviations

The following abbreviations are used in the thesis.

ABA	Achievement-based Assessment
ANOVA	Analysis of Variance
AMO	Assessment and Moderation Officer (NZQA)
BOT	Board of Trustees
CATS	Common Assessment Tasks
CEO	Chief Executive Officer
CBA	Competency-based Assessment
CICAQ	Committee of Inquiry into Curriculum, Assessment and Qualifications
GNVQ	General National Vocational Qualifications (UK)
GRC	Grade-related Criteria
ITO	Industry Training Organisation
LEA	Local Education Authority
LM	Local Moderator
MAP	Moderation Action Plan
MAPSW	Moderation and Assessment Project South West (UK)
NEMP	National Education Monitoring Project
NM	National Moderator
NPA	Nationally Prescribed Activity
NPT	Nationally Prescribed Task
NQF	National Qualifications Framework
NVQ	National Vocational Qualifications (UK)
NZCER	New Zealand Council for Educational Research
NZQA	New Zealand Qualifications Authority
OCEA	Oxford Certificate of Educational Achievement
PPTA	Post Primary Teachers' Association
PTE	Private Training Establishment
PUS	Physics Unit Standard(s)
QA	Qualifications Authority

RL	Regional Moderator
ROL	Record of Learning
ROSBA	Review of School Based Assessment (Queensland)
RPL	Recognition of Prior Learning
SBA	Standards-based Assessment
SES	Socio-economic Status
STAG	Science and Technology Advisory Group
TEPA	Tertiary Entrance Procedures Authority (Queensland)
SC	School Certificate examination
SCOTVEC	Scottish Vocational Education Council
SEC	Secondary Examination Council (UK)
SFC	Sixth Form Certificate
TIC	Teacher in Charge
UB	University Bursary examination
US	Unit Standard(s)

Chapter 1

The Evolution of the National Qualifications Framework

This introductory chapter commences with a brief description of the school administration and curriculum reforms that have occurred in the New Zealand education system since the advent of Tomorrow's Schools in 1989. This sets the scene for a discussion of the paradigm shift in assessment associated with the evolution of the National Qualifications Framework (NQF) and the introduction of assessment against Unit Standards (US) in secondary schools. The implementation of US led to considerable debate that focused on unresolved issues related to the quality management of standards-based assessment. These issues are the foundation for the formulation of the research questions investigated in this thesis. The Level 2 *Physics Unit Standards* (PUS) (NZQA, 1995a) were chosen as a subject specific context within which to address these questions. A final section presents an overview of the structure of the thesis.

1.1 School administration and curriculum reform

Education in New Zealand today has to prepare students for participation in a society which is undergoing rapid social and economic change. High youth unemployment, a raised school leaving age, an increasingly multi-cultural society and an increase in women's participation in the work force have changed the nature of the secondary school population. Social problems, such as teenage pregnancies, drug and alcohol abuse and an increase in violent crime and youth suicide, contribute to the growing complexity and diversity of demands on the education system. In addition, rapidly advancing technological development and the pressure to be increasingly internationally competitive have led to increased emphasis on multi-skilling in the workplace and created a need for schools to deliver programmes aimed at both a broader range and higher level of skills.

In response to these pressures on the education system, the New Zealand Government introduced a series of major reforms in the areas of school administration, curriculum development, and assessment and qualifications. The purpose of these reforms was to modernise:

> ... outdated systems, increase educational opportunities and raise educational standards (Ministry of Education, 1993a: 28).

Proposals for school administration reforms were introduced by the Picot Report (1988) which recommended a devolution of decision making away from the central bureaucracy of the Department of Education. These recommendations led to the introduction of *Tomorrow's Schools* (New Zealand Government, 1988), which was a blueprint for the restructuring of school administration. The governance of schools was devolved to Boards of Trustees and a newly established Ministry of Education was given responsibility for educational policy and curriculum reform.

The structure for curriculum reform was outlined in *The New Zealand Curriculum Framework* (Ministry of Education, 1993a). This document formulated the principles that were to underpin all teaching and learning in New Zealand schools. It identified eight essential generic skills that were to be developed in the contexts of seven essential learning areas. Separate curriculum statements were progressively developed for each of these essential learning areas.

Each curriculum statement sets out aims and objectives for eight progressive levels of achievement, ranging from Level 1 (Year 1) to Level 8 (Year 13). For each subject, individual schools develop programmes or schemes of work that address the achievement objectives at the appropriate levels of the curriculum. The programmes of work for Year 12 and 13 physics courses[1] are based on Levels 7 and 8 of *Physics in the New Zealand Curriculum* (Ministry of Education, 1994a). In addition to meeting the curriculum requirements, physics courses in the senior secondary school have to cover the prescribed content of the recently revised School Certificate (SC), Sixth Form Certificate (SFC) and University Bursary and Scholarship prescriptions (UB/S) (NZQA, 1998).

After initiating the school administration reforms and laying the foundation for the new curriculum framework, the government turned its attention to qualification and assessment reform. Reforms in this area resulted in the establishment of a new NQF.

1.2 The evolution of the NQF

The history of assessment for qualifications in New Zealand can in retrospect be described as a gradual transition from a completely external norm-referenced and summative assessment system to a system that includes internal standards-based assessment (SBA) and incorporates diagnostic and formative functions. This philosophic shift in assessment policy is not unique to New Zealand. Broadfoot (In Peddie and Tuck, 1995: 200) likened similar national assessment changes in the United Kingdom in the last decade to a pantomime in which the:

> Cinderella status of educational assessment among educators has been radically transformed.

She described a fairy tale world in which the:

> ... pumpkin of narrow, norm-referenced and negative assessment

is turned into the:

[1] Physics is not one of the essential learning areas. It is a senior subject that comes under the umbrella of *Science in the New Zealand Curriculum* (Ministry of Education, 1993b).

> ... glass coach of a new assessment paradigm in which the emphasis is on procedures that are individualised, constructive, comprehensive and relevant.

The concept of a paradigm shift is useful to describe the evolution of the New Zealand NQF because it provides a contextual overview and identifies the trends. An assessment paradigm may be defined as a set of interrelated concepts that provide the framework for all aspects of assessment and related decision making. A paradigm shift or 'scientific revolution' occurs when the old paradigm is unable to deal with an outstanding problem (Kuhn, 1962). The focus of the following description of the paradigm shift in assessment in New Zealand is not to provide a complete historical analysis, but to investigate the nature and types of assessment that underpin the succession of qualifications leading to assessment against US. A suitable starting point for this discussion is the origin of the present qualifications. Table 1.1 lists some of the key events in the history of assessment for certification starting with the introduction of SC.

1.2.1 Norm-referenced external exams

The original version of SC was introduced in 1934 at the end of Form 5 (Year 11), as an alternative to the Matriculation exam. SC was regarded as a lower status award until Matriculation was abolished in 1944. The Thomas Report (1944) was a landmark in New Zealand education because it introduced a national core curriculum and recommended a reformed SC as:

> ... a general qualification at the end of compulsory schooling (Woods 1992: 283).

The new SC was introduced in 1946 and rapidly became:

> ... the sole award in Form Five as opposed to the Thomas Committee's intention that it would be a more general alternative to University entrance in Form Six (NZPPTA, 1997a).

Table 1.1: History of assessment for qualifications in New Zealand

Year	Report	Qualifications	Type of assessment
1934		SC was established	Norm-referenced, external and summative examination
1944	Thomas Report	Revision of SC (version two) and introduction of University Entrance	Norm-referenced, internal/ external and summative assessment
1966		Introduction of the Bursary Examination in Form 7	Norm-referenced, external and summative assessment
1968		Introduction of single subject passes for SC	
1969	Education in Change PPTA	Introduction of SFC and Local Certificates	Norm-referenced, internal assessment.
1975			Introduction of scaling for SC
1986	*Learning and Achieving*	Abolition of University Entrance	
1988	Hawke Report		
1989		Achievement-based Assessment Trials	Achievement-based internal assessment
1990	Education Amendment Act	NZQA and NQF established	
	Towards a NQF		
1991	*The National Curriculum of New Zealand*		Standards-based assessment
	ABA in NZ		
1992			Abolition of scaling of SC marks
1994	*Physics in the New Zealand Curriculum*	National Certificate, National Diploma US	Consultation on draft US
1995	*PUS*		PUS Registered
1996	Review of PUS	PUS trial	
	Green Paper: *A Future Qualifications Policy for New Zealand.*		
	PPTA Framework Inquiry report		PUS Revised
2001		Achievement Standards	Standards-based assessment

The exam marks were scaled and:

> ... a pre-determined proportion of candidates were destined to fail in the interest of maintaining 'standards' (Lee and Lee, 1992: 40).

Beeby felt that this was necessary to "maintain the public credibility of the examination (Ibid.) Students passed SC if their total aggregate mark for English and their three best other subjects was 200 or higher (Hughes and Lauder, 1990). Under this system students could get high marks for individual subjects but miss out on the award of SC because of lower performance in English or other school subjects. To address this inequity, the aggregate pass system was changed to a single subject pass system in 1968 (Ibid.). In 1975 scaling of SC results was introduced to maintain longitudinal comparability of the distributions of results for each subject and maintain the subject hierarchy of means (Ibid.).

Concerns about the fairness of maintaining different means for different subjects, equity, and the validity of the underlying assumptions on which scaling is based (St George, 1985; Snook and St George, 1986; NZPPTA, 1997a) led to its abolition in 1992. The introduction of internal assessment for some subjects and internally assessed components for others further changed the nature of SC.

The externally examined UB/S award was introduced in 1966 at the end of Form 7 (Year 13). It is regarded as the "paramount selective device for higher education" (Hearn, 1996b; 17). More recently, some Year 13 subjects became internally assessed or introduced internally assessed components that contributed towards the final UB/S grade.

The SC and UB/S awards fit the characteristics of the psychometric paradigm described by Gipps (1994: 5). This paradigm of assessment developed out of intelligence testing and is based on the assumption that intelligence and scholastic ability are normally distributed and fixed characteristics. Both SC and UB/S are norm-referenced, summative and predominantly externally assessed. Student results are expressed as a single grade or percentage that is regarded as objective and accurate because of a strong emphasis on technical issues such as standardisation, validity and reliability. The primary purpose of this high stakes form of assessment is to generate data for certification. This represents a 'trickle down' form of assessment because results are collected by the NZQA at the end of the year and communicated to teachers and students in the following year. A consequence of this is that the results cannot be used diagnostically or formatively in the current year.

The biggest changes in assessment for qualifications in the senior secondary school have occurred in Year 12. These changes are rooted in the introduction of internal assessment and associated moderation procedures.

1.2.2 The advent of internal assessment

After the release of the Thomas Report, University Entrance (UE) was introduced at the end of the sixth form year (Year 12). This new qualification was awarded on the basis of school-based, norm-referenced assessment. Schools were able to accredit students with UE, but had to meet the requirement that a proportion of the students, who were not accredited, passed the external exam each year. If a school accredited too many students and consequently did not achieve the required pass rate in the external examinations, the school's authority to accredit could be withdrawn (Woods 1992). This moderation procedure provided a check on the consistency of assessment standards between schools. It also provided a basis for comparability of results attained by students from different schools.

In the 1960's the combined effects of rapid post-war population growth, the raising of the school leaving age to 15 and the increasing retention rate, changed the nature of the sixth form student population. An increasing proportion of sixth form students did not intend to proceed to university study and found that UE was either too difficult or inappropriate for them. Elley and

Livingstone (1972:14) remarked that since less than 15% of sixth form students proceeded to university study:

> The University Entrance Examination is apparently not meeting the needs of many thousands of sixth formers.

By the late 1960's the universities, the Department of Education, and the New Zealand Post Primary Teachers' Association (NZPPTA) pushed for the establishment of an alternative sixth form qualification which was not narrowly focused on academic subjects and enabled the development of local school-based courses. This culminated in the introduction of SFC in 1969 (NZPPTA, 1997a).

The internally assessed SFC is norm-referenced. For each sixth form subject offered by a school, grades ranging from 1 (high achievement) to 9 (low achievement) are awarded on the basis of a ranked list of internal marks. To maintain comparability between schools, the internal school marks are scaled to a distribution of marks derived from the students' performance in the previous year's SC examinations. The award of SFC existed alongside the more prestigious UE qualification, as a secondary and less prominent educational qualification and was overshadowed to the extent that it did not gain widespread recognition.

In 1971, the NZPPTA (1971) signalled that the presence of two qualifications in the sixth form was unworkable. Dissatisfaction with the appropriateness of UE gained momentum and resulted in a threat by the NZPPTA for members to boycott the UE exam in 1983. The new Labour Government supported the NZPPTA's policy and UE was eventually abolished from 1986 (Hearn, 1996b; Lee and Lee, 1992).

The abolition of UE placed more emphasis on SFC and its award and moderation procedures. There were several problematic issues. Since the SC examination acted as a moderating test for the award of SFC grades, it established a hierarchy of subjects that limited the number of high grades that could be awarded to students in some subjects. This was seen to be unfair because the allocation of grades was norm-referenced and did not necessarily reflect students' performance relative to the course objectives in these subjects. Since the distribution of grades was pre-determined at the start of the year, there was no allowance for changes in the overall performance of the cohort during the sixth form year.

The increase in the number of internally assessed Year 11 subjects reduced the statistical validity of the scaling procedures in Year 12. In the case of internally assessed SC science, the distribution of school results was generated by student performance in a reference test that was based on Year 9 and 10 work. This internally assessed Year 11 mark was used the following year to contribute to the Year 12 grade distribution. The combination of these factors led to a growing dissatisfaction of teachers and students with this form of statistical moderation.

To address these issues, the Labour government commissioned a series of research reports on post-compulsory curriculum, assessment and qualifications. The first major report, *Learning and Achieving* (Department of Education, 1986), was produced by the Committee of Inquiry into Curriculum, Assessment and Qualifications (CICAQ). The committee considered public submissions on the discussion document *Assessment and Awards in the Senior Secondary School.* The report's recommendations that related to sixth form assessment were that:

> ... in the moderation of assessment, each subject be considered independently of others and without reference to achievement at other levels (15),

and that:

> ... urgency be given, to the setting up of trials to investigate alternative moderation procedures for sixth form subjects (18).

The report suggested SBA as a possible form of assessment to achieve these recommendations. This led to investigations into SBA.

1.2.3 Investigations into SBA

The Board of Studies (1988) defined SBA as:

> ... assessment of learner achievement against specified, published standards.

There are two types of standards-based assessment. Sass and Wagner (1992: 13) explained that:

> The term standards-based assessment is used to encompass both the competency-based approach of mastery learning appropriate to the vocational field and to the achievement–based approach typically used in general education.

Peddie (1992: 27) defined achievement-based assessment (ABA) as:

> ... assessment in which a number of progressively more demanding standards are used, and in which all learner achievement is reported as a letter or number grade.

Competency-based assessment is:

> ...assessment where we set a particular standard which candidates must reach, if they are to be judged as competent and therefore receive credit for the unit of learning (Ibid.: 24).

The vocational sector started to develop the competency-based approach, whereas in the secondary school sector, ABA was introduced in 1987, to investigate whether it could be used to establish SFC as a stand-alone qualification. The form of ABA trialled used five levels to report students' achievement.

The descriptions of the achievement required for these five progressive grade levels were called Grade Related Criteria (GRC). The report *Achievement-Based Assessment in New Zealand* (NZQA, 1991a) contains the GRC that were developed for each sixth form subject involved in the ABA trials. The GRC for physics and an associated marking schedule for a specific assessment task are contained in Appendix 1.

The ABA trials involved about a third of all New Zealand secondary schools (NZQA, 1991a) and were accompanied by a comprehensive programme of professional development for teachers. In 1989 and 1990, about 1700 sixth form teachers were trained in the use of ABA for geography, biology and physical education.

ABA was school-based and the assessment activities were designed and administered by teachers. It is a trickle up form of assessment where schools collect the assessment data that is sent to NZQA at the end of the year.

ABA is a form of criterion-referenced assessment and can be used diagnostically to assess students' prior learning before commencing instruction and to "diagnose learning difficulties during instruction" (Linn and Gronlund, 1995: 14). At the class level the outcomes of ABA assist teachers in programme planning and identifying learning needs, so that 'appropriate remedial' or accelerated instruction may be provided (Irwin, 1994: 47). At the student level it targets:

> ... difficulties that students may be having, to determine their precise nature and scope, and to plan further learning activities to meet the needs of those students (Ministry of Education, 1994b: 8).

ABA also has formative and ipsative functions. It can be used to "monitor learning progress during instruction" (Linn and Gronlund, 1995: 14) and enables students to track and plan their own progress by monitoring the extent to which the learning outcomes stated in the staircases of descriptors have been achieved. The results provide the teacher and student with information on what further learning needs to occur to achieve at a higher level.

Following the recommendations made by the Hawke report (1988), the 1990 Education Amendment Act established the New Zealand Qualifications Authority (NZQA) and charged it with the responsibilities of running the existing norm-referenced qualifications system, implementing assessment and qualification reforms and establishing a NQF.

At the end of 1990 the NZQA issued the discussion document *Towards a National Qualifications Framework* (NZQA, 1990). The public submissions expressed support for SBA and the development of units of learning that incorporated ABA (NZQA, 1996a). The consultation document *Designing the Framework* (NZQA, 1991b) advocated the use of both competency- and achievement-based assessment and recommended that:

> ... the most appropriate form of standards-based assessment should be selected for each area of learning (NZQA, 1991b: 58).

It concluded that:

> Competency-based assessment is most suitable for areas of learning in which learning outcomes can be described in terms of a discrete skill that can be performed to a defined standard.

and that:

> ABA is better suited to general subjects in which attributes and skills are not measurable in quantitative terms, but rely on qualitative judgement (NZQA, 1991b: 58).

In July 1992, the NZPPTA placed a moratorium on members' involvement in curriculum and assessment reforms as part of industrial action against the proposed introduction of bulk funding of teacher salaries (NZPPTA, 1992). This effectively stopped the development of units of learning which incorporated ABA. The moratorium was suspended in December 1992 (NZPPTA, 1993). In the intervening period the vocational sector continued with the development of competency-based units of learning based on the Scottish SCOTVEC model. This overtook the original ABA pilots.

To bridge the vocational/academic divide, it was decided that there would be no separate academic and vocational qualifications. The consequence of this decision was that the Framework would use the same form of assessment for all registered qualifications and that qualifications would be linked through common units of learning (NZQA, 1991b). Since the vocational competency-based model was well developed at this stage, it was adopted as the common form of assessment for the New Zealand Qualifications Framework.

1.2.4 The NQF and US

The NQF was set up under Section 53 of the 1990 Education Amendment Act. It consists of eight levels. Levels 1-3 correspond to Years 11-13 of the senior secondary school. Most school students will earn credits towards the National Certificate of Educational Achievement. The NQF also offers National Certificates or Diplomas in specific vocational areas. Study towards these qualifications can be commenced at secondary school and completed beyond school. This is referred to as the *seamlessness* of qualifications.

The building block for achieving these qualifications is the US. A US is:

> ... the standard against which evidence collected on learner performance is judged (NZQA, 1995b: 19).

All registered US are assigned to a level on the NQF and are made up of specific learning outcomes called elements which:

> ... describe the competencies which must be demonstrated for the achievement of the US (NZQA, 1995b: 19).

Assessment against US is internal and competency-based. Assessment activities are designed to give students the opportunity to present evidence that they have met the various performance criteria and consequently the learning outcomes of the elements that make up a US. To receive credit for a US, students must meet all of the performance criteria for each of the elements of the US.

When a student is first assessed against a US, the results indicate exactly which performance criteria have not been achieved and direct further learning. This gives the assessment a diagnostic, ipsative and formative component. The assessment is summative in the sense that, at the end of the year it leads to a Record of Learning (ROL). This is issued annually by NZQA, and records the US credits that have been achieved by the student. The assessment does not lead to a global grade like SFC.

The development of the US for each subject was overseen by NZQA appointed advisory groups. The Science and Technology Advisory Group (STAG) was established in 1995, and the Level 1-3 Physics Unit Standards (PUS) were registered on the New Zealand NQF on 22 December 1995. The title matrix for the PUS is contained in Appendix 2. To illustrate the practice of assessment, an example of a PUS and an associated assessment activity and marking schedule have been included in Appendices 3 and 4.

In 1995, 114 physics teachers received three days' training in standards-based assessment. The PUS were trialled in 1996 by 122 secondary schools. David Hood (1996), the NZQA chief executive stated that in 1996, at least a third of senior secondary school students were assessed against US in one or more subjects and 82% of schools surveyed were using the Framework.

1.3 The US debate and unresolved issues

The introduction of assessment against US in the senior secondary school resulted in widespread public and academic debate. Contributors to the debate included teachers, students, principals, parents, and academics. The debate prompted several reviews and investigations into the Framework. These included the *Te Tiro Hou* Qualifications Framework Inquiry (NZPPTA, 1997a), The New Zealand Institute of Physics survey (NZIP, 1996), the Education Forum-commissioned review of the Qualifications Framework

(Smithers, 1997) and the NZQA report on the trials (NZQA, 1997a). The unresolved issues identified by the debate and reviews included:

1. The question of recognition of excellence and how to incorporate this into the Framework

The ROL issued by NZQA each year, records whether students have or have not achieved the US but does not indicate the level of achievement, nor the number of times the student was reassessed. It is argued that this approach does not recognise excellence (Sinclair, 1997; Austin 1996; NZPPTA, 1997a; NZIP, 1996; Rawson, 1997). Salter and Hayden (1996) found in their trial of the Biology US that:

> ... some students constantly achieved what is considered a high standard of work while others performed only satisfactory work, yet both groups of students passed US with no distinction between the quality of their work.

The consequence of the lack of recognition of excellence is that there is no incentive for students to perform at a higher level than the level of competency set by the standard.

2. The impact of assessment against US on student motivation

Critics of assessment against US argued that the lack of recognition of excellence has a negative effect on the motivation of able students. The PPTA review found that it was very difficult for less able students to get credit on the Framework and that this had an adverse effect on their motivation and enjoyment of the subject (NZPPTA, 1997a). Vlaardingerbroek (1996) who interviewed the Heads of Science Departments at 18 Palmerston North and Dunedin schools supported this view. One HOD worried that:

> ... with all-or-nothing assessment, they're going to be a Grim Reaper for many kids.

3. *Can US adequately and unambiguously describe a standard?*

Coogan (1996: 14) stated that "transparency is a key principle of competency-based assessment" and that US:

> ... must be so specific and unambiguous that both learners and assessors can interpret them consistently.

Smithers (1997: 78) doubted that US could fulfil this requirement and felt that:

> Experience has shown that US cannot be stated with the precision necessary to ensure the fairness, consistency and validity of assessment, and therefore the attempt to use them as the common currency for a qualification structure should be abandoned.

There is considerable debate about whether US can adequately define a standard, or whether the standard resides in the assessment task, the judgement statements in the assessment schedule, the moderation process, the collective experience and socialisation of the teachers using the system or a combination of these (Irwin, Elley and Hall, 1995; Batchelor, 1996; Irwin, 1995). David Nicholson in Smithers (1997: 42) used Fig 1.1 to illustrate that the standard is defined by a process of triangulation between the registered standard, the assessment activity and schedule, and the moderation process.

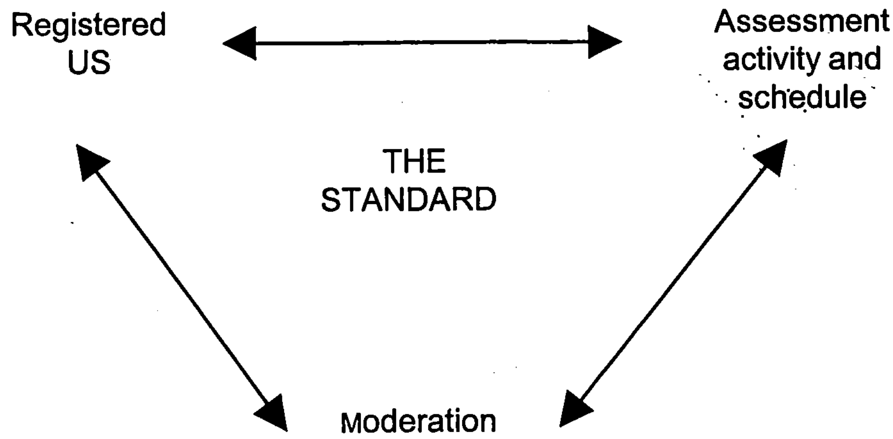


Fig 1.1: The location of the standard

4. Can US be used to validly assess all subjects in the school curriculum?

The NQF uses assessment against US for both vocational and academic subjects. Critics feel that US are a valid way of assessing vocational and skills-based subjects but cannot be used to assess the higher order thinking skills which characterise academic subjects and that the fundamental weakness of the Framework has been to force vocational and academic courses into the same assessment model. They argued that trying to reduce everything to competency-based assessment is not feasible on the scale of the framework (NZPPTA, 1997a; Godfrey in Sinclair, 1997; Marshall, 1994; Irwin, 1995).

Critics feel that assessment against US is unsuitable for measuring students' understanding of school subjects which involve large bodies of knowledge (history, biology, economics etc.) and skills such as problem solving, writing and critical thinking (Elley, 1994). These problems have been described for mathematics (Neyland, 1994), history (Childs, 1995), science (Austin, 1996), and physics (Batchelor, 1996; the New Zealand Institute of Physics, 1996). Salter and Hayden (1996) said that:

> ... it was doubtful whether everything that occurs within a classroom can be assessed by criterion-based methods and there is a danger in using US as the only means for recognising how a student performs at school.

Carter (1996) concluded:

> ... that our collective experience to date has brought us to the inescapable conclusion that the system of US assessment is inherently flawed as a method of measuring what students can do in the context of traditional school subjects.

5. The impact of assessment against US on teaching programmes.

Critics have argued that not all learning can be assessed by competency-based assessment. They worry that the Framework will reduce education to a collection of US being defined solely in terms of discrete, observable and measurable behaviours (Marshall 1994; Irwin, 1995; Salter and Hayden, 1996). Smithers (1997: 44) agreed and expressed the caution that:

> ... the essence of subjects almost seems to disappear in the attempt to express them as numerous performance criteria.

Along these lines, Sinclair (1997: 2) expressed concern that assessment against US "can stamp out creative flair". Because of this spiral to specificity, assessment against US can start to dominate the design of teaching programmes and lead to the fragmentation of teaching and learning (NZPPTA, 1997a).

6. Is the moderation system achieving acceptable comparability between schools?

Smithers (1997) claimed that the expectation that the moderation systems set up by the NZQA can achieve consistency of assessment against US across nearly 450 secondary schools is looking dauntingly difficult to achieve. Doubts about the consistency and manageability of the moderation systems have been raised. (Irwin, 1995; Mann, 1997; Morris, 1996; NZPPTA, 1997a; Smithers, 1997). Finch (1994) posed the question:

> How do we moderate internal SBA between a school in Whangarei and one in Invercargill?

Salter and Hayden (1996) felt that because there is a large number of moderators involved in moderating assessment tasks for different schools, there is a potential for inconsistencies in assessor judgements to develop. Carter (1996) experienced varying standards between local moderators and considered this to be evidence that there must be "appreciable variations on a national scale" which render the system unreliable.

Reassessment activities are not moderated and Salter and Hayden (1996) argued that if reassessment is casual, as in the case of verbal reassessments or resubmission of scripts during class time, it can require a different standard of student performance than the original formal assessment. Chamberlain (1996) felt that there is a temptation for teachers to be less than rigorous in order to enhance the school's academic record by getting lots of students through their US.

Rawson (1997: 4) raised the issue of school-based moderation and the unknown consistency between teachers within a department. This aspect of consistency cannot be controlled by the moderation system, but the accreditation process requires each provider to have procedures for internal moderation in place.

7. *Issues related to the practicality and manageability of assessment against US*

The implementation of assessment against US raises issues of manageability and practicality such as workload, dual assessment and resourcing.

The increase in teacher workload associated with assessment against US is seen by critics as a major drawback (NZPPTA, 1997a; Rawson, 1997). Retired Christ's College Principal, Dr Maxwell Rosser (1996) cautioned that there is a danger that:

> ... so much time will be allocated to these assessment procedures that teaching time and, therefore, education itself will suffer.

The increase in workload can be attributed to:

- **The preparation of assessment activities and schedules**

  Salter and Hayden (1996) took part in the 1996 trial of the Biology US and estimated that about 15 hours were spent on each US before it could be offered to students and that it took about 100 hours to prepare each course.

- **Administration relating to moderation, record keeping and communication with NZQA**

  Critics argue that keeping student records, submitting assessment activities for moderation, keeping up-to-date with communications from NZQA and submitting results leads to an unacceptable workload, which is further compounded by reassessment (Carter, 1996). Salter and Hayden (1996) said that reassessment took up teaching time and that lunch hours were used to reassess students. This increase in workload will be exacerbated threefold when US are implemented at three levels.

- **Dual assessment.**

  In 1996 Education Minister Wyatt Creech announced that at level 2, schools could offer SFC or US or a combination of both. Sewell (1996) and Wilson (1996) argued that part of the increase in workload was associated with dual assessment while SC, SFC and UB/S remain in their present form. This should reduce when SFC is abolished. It is argued that the increase in workload due to dual assessment is unsustainable for any lengthy period (Hearn, 1996a).

Teachers argued that the resources for the implementation of the Framework have not been sufficient (Braun, 1996; Wilson, 1996). Suggestions were made that the government should compensate teachers for the increased workload by providing time allowances to implement US (Batchelor, 1996; Rawson 1997), and funding for ancillary staffing to process student data (Rawson, 1997). Rawson (1997:4) listed other manageability issues raised by trial teachers, such as, the storage of student portfolios, the cost to students of registering credit on the Framework and the need for item banks to avoid teachers re-inventing the wheel.

8. Political uncertainty about the final structure of the New Zealand Qualifications Framework and the future of US and the present qualifications

The results of a 1996 survey of secondary schools (Morris 1996) showed that 55% of schools surveyed supported a dual pathway on the Framework and proposed that the norm-referenced and standards-based qualifications systems be integrated. Gernhoefer (1996) was positive that such a change could be incorporated in the Framework. One of the recommendations put to the 1997 annual PPTA conference was that:

> PPTA supports the development of SC and Bursary as awards using standards-based assessment, earning credit on a modified Qualifications Framework and that SFC be abolished once a modified Qualifications Framework is in place. (NZPPTA, 1997b: 2)

The history of qualifications in New Zealand indicates that whenever there are two qualifications at the same level of schooling, the qualification that is perceived as more prestigious will overshadow the other. This could be the case if the traditional secondary school qualification structure is allowed to remain concurrently alongside US (Vlaardingerbroek, 1996).

Braun (1996) reporting back on an Otago and Southland Secondary Principals' Association conference on the Framework stated that the majority of principals did not support a dual pathway proposal and wanted to retain bursary as an endpoint qualification. Green Bay High School principal, Karen Sewell, (1996) said that the concept of a dual pathway had:

> ... little educational value and would undermine the NQF.

In 1997 the Minister of Education summed up the hiatus in New Zealand's qualifications policy:

> New qualifications, based on US, have gained wide acceptance in areas such as industry training and non-traditional subjects in secondary schools. However, many people in education circles have raised significant concerns about the policy of implementing a system based exclusively on US across all subjects and educational providers. There are both fundamental objections to the system and technical concerns about its operation, notably the extra workload assessment and moderation puts on classroom teachers (Ministry of Education, 1997: 4).

To resolve this hiatus, a Green Paper on a future qualifications policy for New Zealand was released on the 5 June 1997 which invited public submissions on proposals relating to the implementation of the NQF. The specific proposals relating to the senior secondary school were that:

- the development of US and qualifications assessed against US should continue and incorporate a scale of nationally recognised excellence (Ministry of Education, 1997: 8)
- that national school examinations may be registered as qualifications on the NQF and yield credit towards the National Certificate in Educational Achievement (Ibid.: 26)
- that once the National Certificate is fully operational across all school subjects, it could replace the current SFC (Ibid.: 27).

The White Paper that followed led to the announcement of *Achievement 2001* (Ministry of Education, 1998) which is a unified system of national qualifications that incorporates examinations, achievement standards and US. This will be phased in commencing in 2001.

1.4 Conclusion

This chapter placed the evolution of the New Zealand NQF in the context of a paradigm shift from a psychometric norm-referenced paradigm to a broader model of educational assessment. This paradigm shift is neither complete nor universally endorsed. Significant unresolved tensions remain between the political agendas of the New Right and those of professional educators. Codd, McAlpine and Poskitt (1995: 32) argue that we have arrived at a Janus type policy that is shaped by:

> ... conflicting political imperatives in which there is a fundamental tension between assessment that aims to improve learning in relation to the needs and abilities of individual learners, and assessment that constitutes a mechanism for centralised control and accountability.

These tensions are reflected in the unresolved issues identified by the public debate and the commissioned research discussed in this chapter.

The unresolved issues relate to the quality management system set up by NZQA and need to be addressed by detailed longitudinal research into the validity, reliability and manageability of US-based assessment in the senior secondary school and its impact on individual subjects, teachers and students. Since the author is the National Moderator and one of the writers of the PUS, physics was selected as a subject specific context for the present research. The main question investigated is whether assessment against the PUS and the associated moderation, provide a valid, reliable and manageable method for assessing the achievement objectives of *Physics in the New Zealand Curriculum* (Ministry of Education, 1994a). This question is broken down into a number of main and subsidiary research questions that are outlined in Chapter 2.

1.5 Outline of the structure of the thesis

Chapter 1 described the evolution of the NQF leading to the 1996 PUS trial in the context of a paradigm shift away from psychometric assessment towards a broader model of educational assessment. For assessment against US to gain acceptance as a viable alternative to the established system, it is important that there is public confidence in the quality management system set up by NZQA. The PUS are nominated as a subject specific context to investigate this issue in detail.

Chapter 2 explores issues of quality management of SBA. It draws the distinction between quality control and quality assurance. It introduces validity, reliability and manageability as traditional quality indicators of assessment and reconceptualises these indicators for the evaluation of standards-based assessment. It contains a literature review of moderation and looks at the New Zealand and overseas experience of moderation of SBA. The Moderation Action Plan for science and the role definitions of the National, Regional and Local Moderators are described in detail. A concluding section outlines the research questions that aim to investigate the unresolved issues outlined in Chapter 1 in the context of the validity, reliability and manageability of assessment against the PUS.

Chapter 3 commences with an outline of the research design. It describes the subjects and explains the sampling techniques used to select them. The research instruments are described in detail and a final section describes the quantitative and qualitative techniques used to analyse and report the data gathered by the research.

Chapter 4 reports on the validity of Year 12 physics assessment. Aspects of validity considered include curriculum fidelity, content, concurrent validity and the validity of reporting of assessment against the Level 2 PUS. It also discusses consequential validity of the PUS by examining the impact of assessment on student learning and class teaching. Comparisons are made between assessment for SFC and US.

Chapter 5 reports on a three-year investigation into the Level 2 PUS moderation system. This commences with the trial in 1996 to wider implementation in 1998. The chapter presents teachers' and moderators' opinions on the effectiveness of the various components of the Moderation Action Plan for assessment against the PUS. It further reports on the outcomes of annual moderator and end-point assessor judgement agreement trials.

Chapter 6 discusses issues related to the implementation and manageability of assessment against the level 2 PUS. It discusses the workload, resourcing and the school-based implementation issues related to assessment against the Level 2 PUS. Annual student and teacher questionnaires and longitudinal case studies of five Canterbury schools collected the data on which this discussion is based. The case studies traced each school's involvement with the PUS over a three-year period from 1996-98.

Chapter 7 discusses the key research findings in relation to the literature and makes recommendations for improving the process of assessment against the PUS and the accompanying moderation procedures. The recommendations are specific to the PUS, the moderation system, assessors and NZQA. A final section suggests possible avenues for future study in the field of quality assurance of standards-based assessment.

Chapter 8 discusses the implications of the research for assessment against US for conventional school subjects in general. The discussion identifies the outstanding issues that need to be resolved before assessment against US is officially implemented across all conventional school subjects at levels 1-3 of the NQF and makes some recommendations for the system as a whole.

Chapter 2

Quality Management of Standards-based Assessment

This chapter reviews the literature on the quality management of SBA. It identifies the key components of quality in assessment and draws the distinction between quality control and assurance. Validity, reliability and manageability are introduced as traditional key quality indicators of assessment. These indicators are reinterpreted to apply to the evaluation of assessment against the Level 2 PUS. The extent to which the tensions between these quality indicators of SBA can be resolved depends on the effectiveness of the accompanying moderation procedures. This chapter examines New Zealand and overseas research into moderation and describes a wide range of moderation techniques. It focuses on the quality assurance procedures incorporated in the Science Moderation Action Plan (MAP) that prescribes the moderation procedures for assessment against the Level 1-4 PUS on the NQF. A final section states the research questions.

2.1 Key indicators of quality management of assessment

Any system of assessment for national qualifications has to be of high quality to be publicly acceptable and credible. The concept of quality is generally defined in relation to products and services. In the context of education, the product is:

> ... the new skills and knowledge acquired by the educated or trained person

and the service is:

> ... the provision of an environment that enables the new skills and knowledge to be acquired (NZQA, 1993: 6).

Harlen (1994: 13) defined quality in assessment as:

> ... the provision of information of the highest validity and optimum reliability suited to a particular purpose and context.

In relation to the quality management of assessment, Harlen (1994) identified two main categories: quality control and quality assurance.

The emphasis in quality control is on post-assessment procedures that adjust the outcomes of assessment in order to improve fairness for the groups and individuals being assessed. This aspect of quality management is closely associated with the norm-referenced assessment paradigm outlined in Chapter 1 and does not impact on the task design and preparation stages of assessment. An example of a quality control procedure is the post-assessment scaling of examination marks.

The emphasis in quality assurance is on optimising each step of the assessment process to arrive at fair assessment for groups and individuals (Harlen, 1994: 16). The concept of quality assurance fits into the educational assessment paradigm outlined in Chapter 1 and is more relevant to school-based assessment against US, where teachers are involved at all stages of the assessment process.

The quality assurance process for assessment against US needs to address the three key quality indicators that are central to evaluating any form of assessment. These indicators are validity, reliability and manageability (Thorndike et. al., 1991: 91). However, Hearn (1996b: 33) observed that:

> The concepts, language and methodology used by psychometricians to measure the quality of standardised norm-referenced testing were developed for that model and cannot be applied unquestioningly to new forms of assessment.

The following sections reinterpret how each of these quality indicators may be used to evaluate assessment against US.

2.1.1 Validity of SBA

Gipps (1994: 58) defines validity as the:

... extent to which a test measures what it was designed to measure. Aspects of validity of assessment that have traditionally been used include predictive validity, concurrent validity, construct validity and content validity. In addition, the more integrated notions of curriculum fidelity (Gipps, 1994), consequential validity (Messick, 1989) and systemic validity (Frederiksen and Collins, 1989) are relevant for evaluating assessment against the PUS.

1. Predictive validity

Predictive validity of an assessment task is a measure of the extent to which student performance in the task can be used to predict performance on future assessment tasks. The predictive validity of assessment against the Level 2 PUS can be established by comparing the achievement of a group of students on Level 2 PUS assessments with their performance in the UB/S physics examination the following year. Hearn (1996b: 37) questioned the usefulness of the concept of predictive validity. She argued that student performance depends on the context and conditions under which assessment takes place and deemed it inappropriate to generalise to other contexts. For instance it is inappropriate to generalise from an internally assessed 30 minute practical Year 12 PUS task to performance in a three hour external Year 13 examination. For this reason this aspect of validity has not been investigated in the present research.

2. Concurrent validity

The concurrent validity of an assessment activity is a measure of how well the assessment results for a group of students correlate with the results of a related assessment of the same skills, knowledge and understanding (Hearn, 1996 b; Kempa, 1986). The concurrent validity of assessment against the Level 2 physics PUS can be determined by comparing the distribution of the total number of Level 2 physics credits achieved by a group of students with the distribution of their SFC physics grades at schools that practise dual assessment. This is, of course, based on the assumption that SFC grades are valid.

3. Construct validity

The construct validity of an assessment activity is a measure of whether the assessment is an adequate measure of the constructs or underlying skills that are required to respond to the assessment items. Messick (1989: 7) identified two major threats to construct validity namely:

> ... construct under-representation, that is, the test is too narrow and fails to include important dimensions of facets of the construct and construct- irrelevant variance, that is, the test contains excess irrelevant variance, making items or tasks easier or harder for some respondents in a manner irrelevant to the interpreted construct.

This notion has particular relevance to the nature of the performance criteria of the PUS. It is essential that these criteria are relevant to the learning outcomes stated in the elements.

4. Content validity

Content validity is a measure of the extent to which an assessment samples the content and objectives of the course or curriculum to which it relates. Gronlund (1973: 47) claims that:

> ... during the construction and use of criterion-referenced tests we are concerned primarily with content validity.

A report of the Secondary Examination Council (SEC, 1986: 5) in the UK stated that for assessment against criteria:

> ...to be valid, the criteria employed must emerge from the curriculum, be constructed in collaboration with teachers and be subject to regular revision.

For the Level 2 PUS to have satisfactory content validity they must accurately reflect Level 7 of *Physics in the New Zealand Curriculum* (Ministry of Education, 1994a).

5. Curriculum fidelity

Hearn (1996b: 38) warned that the breakdown of validity into separate categories could result in an atomised approach to test validation that can lead to aspects of validity being considered in isolation or ignored by teachers.

In addition, the problem with the technical nature of validity and its expanding definitions is that the concept can become too cumbersome for classroom teachers to apply in evaluating assessment design. This is especially problematic for assessment against US where teachers design high stakes internal assessment activities for national qualifications. To simplify the notion of validity and make it more accessible for teachers to use in task design, Gipps (1994) suggested 'curriculum fidelity' as an alternative quality criterion. This incorporates the notions of content and construct validity. For an assessment to have high curriculum fidelity it requires:

> ... that the construct, domain or curriculum is well specified and that there is broad coverage of the curriculum (if not each domain) in the assessment (Ibid.: 174).

This notion is useful for the validation of assessment against the PUS.

6. Consequential validity

Messick (1989: 19) argued that validity should be seen as a unitary concept that includes "the social consequences of the testing". Messick's value implications of the consequences of assessment relate to the notion of systemic validity elaborated by Frederiksen and Collins (1989: 27):

> A systemically valid test is one that induces in the education system curricular and instructional changes that foster the development of the cognitive skills that the test is designed to measure.

An investigation of the social impact of assessment against US should be considered as part of the quality assurance process and should include a consideration of the impact of this form of assessment on teacher enthusiasm, students' learning, students' motivation and students' enjoyment of the Year 12 physics course.

More recently Crooks, Kane and Cohen (1996: 1) proposed a validity evaluation model, which depicts assessment as a chain of eight linked stages: administration, scoring, aggregation, generalisation, extrapolation, evaluation, decision and impact. Threats to validity are associated with each link and overall validity is limited by the weakest link. The links in this model provide useful guidance on what aspects of assessment against the PUS to include in the evaluation.

Based on the above discussion, an investigation into the validity of assessment against the PUS needs to consider concurrent validity, curriculum fidelity (including content and construct validity), validity of the reporting process and consequential validity. The questions that need to be addressed are:

- Do the PUS adequately represent the achievement objectives of *Physics in the New Zealand Curriculum*?
- Do the assessment activities used allow students to present evidence that they have attained the performance criteria and consequently the elements of the PUS?
- Are the number of credits allocated to each PUS proportional to the proportion of a provider's programme allocated to the development of the unit content?
- How valid are US credits for describing student achievement in physics and how it can be improved?
- What is the impact of assessment against the PUS on teaching and learning?

2.1.2 Reliability of SBA

Reliability is a measure of the accuracy or precision of an assessment and the extent to which the scores are reproducible. Reliability is vital for assessment for national qualifications because it has a direct impact on public confidence and credibility. Shorrocks, Daniel, Frobisher, Nelson, Waterson and Bell (1992: 32) warn that:

> Traditional reliability measures, based on correlation techniques, are likely to be misleading, since they rest on the assumption of high levels of discrimination between pupils and wide variation in scoring.

Therefore, traditional methods for investigating reliability such as the test/re-test or split half reliability techniques are not useful for SBA and the concept of reliability needs to be reconceptualised.

Coogan (1996: 112) argued that reliability of assessment against US is concerned with whether candidates might have received:

> ... a different judgement in another time, in another place or with a different assessor.

An assessment is considered reliable to the extent that different assessors using the same evidence make comparable judgements or that the same assessor over time makes comparable judgements. For Gipps (1994:174), comparability:

> ... is achieved through consistency of approach to the assessment by teachers; a common understanding of assessment criteria; and that performance is evaluated fairly, that is, according to the same rubric by all markers.

The reliability of assessment against the PUS depends on the comparability of assessor judgements between different providers, which in turn depends on the:

- quality and specificity of the US
- consistency of interpretation of the US by assessors
- formulation of judgement statements which allow consistent decisions about whether or not the performance criteria have been attained
- consistency of end-point assessor judgements

- internal consistency between assessors within the same provider
- consistency of moderator decisions
- assessor's expertise and experience in SBA
- effectiveness of the moderation of end-point assessor judgements.

2.1.3 Manageability and workload issues

In addition to aiming for high validity and reliability, any new system of assessment has to be manageable for teachers to implement.

> Manageability relates to the practical feasibility of achieving validity and reliability in such a way that the assessment system does not impose undue stress, professional burden or workload on those implementing it (Gilmore in NZPPTA, 1997a: 39).

Based on his monitoring of the GNVQ experience in the UK and SCOTVEC in Scotland, Coogan (1996: 36) highlighted manageability as a key barrier to the implementation of a full SBA system. He found that:

> Teachers especially, but also educational administrators in central agencies, increasingly view a pure model of competency-based assessment as an ideal which is difficult to attain in schools given the realities of the classroom and the limits on resources.

He identified the following factors that adversely affected manageability:

1. Content coverage

Because the elements cover all the curriculum content, there is a tendency to attempt to assess all of the course outcomes instead of sampling. There is a danger that courses can place too much emphasis on meeting assessment requirements and become "assessment driven".

> This can place impossible demands on both teachers and students (Ibid.: 37).

2. Administration overload

Recording of student achievement, maintaining portfolios and excessive clerical demands can be tedious and time consuming because of the:

> ... sheer volume of information awarding bodies have required centres to record (Ibid.: 54).

3. Class size

Large class sizes can make it difficult for teachers to make judgements about a wide range of criteria for a large number of students. The assessment demands placed on teachers of large classes are a "key determinant of manageability" (Ibid.: 49).

4. Nationally prescribed tasks

The introduction of nationally prescribed tasks that aim to increase inter-assessor reliability was seen by teachers as a major contributor to workload, a "waste of time" and "unnecessary" (Ibid.: 54).

These overseas concerns about workload and resources were echoed by New Zealand secondary school teachers when the US trials were held in various school subjects and led to the NZPPTA freeze on members' participation in new Framework development in 1996. The unresolved manageability issues that need to be properly addressed if the PUS are to be implemented successfully include teacher workload, the amount of time spent on assessment and administration, the suitability and availability of resources, professional development, and the impact of assessment on students and teachers.

2.1.4 Tensions between validity, reliability and manageability

Ideally, assessment against US should be highly valid, of optimal reliability and easy to manage for teachers. It is unlikely that any assessment can satisfy these three criteria completely and simultaneously. Consequently the quality of any assessment is a compromise between these three quality indicators. This raises the question of the relative emphases that the quality assurance process should place on each indicator. Burke and Jessup (1990: 195) consider reliability to be less important than validity in SBA. They argue that:

> reliability is vital in any norm-referenced system because by definition it is concerned with comparing one individual with another but in a criterion-referenced assessment the intention is very different. Once external explicit criteria have been established, there is an external reference point for assessment. The essential question of validity centres on comparing the judgements made (by the assessors) with the criteria and not between different assessors or assessments. In these circumstances reliability is not an issue.

Gipps (1994) concurred that the emphasis in SBA is on validity rather than technical reliability. Thorndike et. al. (1991: 91) commented that:

> Validity is the absolutely essential quality for a test to have, but in a sense, reliability is a necessary precondition for validity. Test scores must be at least moderately reliable before they can have any validity, but a reliable test may be devoid of validity for the application we have in mind.

Both reliability and validity have to be considered in relation to the contexts and purposes of assessment. In regular school-based assessment the emphasis is on validity, whereas in assessment for qualifications the emphasis shifts towards reliability. In the case of assessment against US, tensions between validity and reliability occur because assessment is used simultaneously at the school level to support teaching and learning and at the national level for the award of the National Certificate. Central to these tensions is the debate over whether the perceived high validity of assessment against US is achieved at the expense of reliability. This could lead to unacceptable variations between assessors and adversely affect public credibility. In addition there are tensions between validity and reliability, and manageability. If concerns about manageability issues lead to inadequate sampling of course outcomes or inadequate moderation, validity and reliability may be adversely affected. These tensions may be illustrated by the triangle in Fig 2.1.

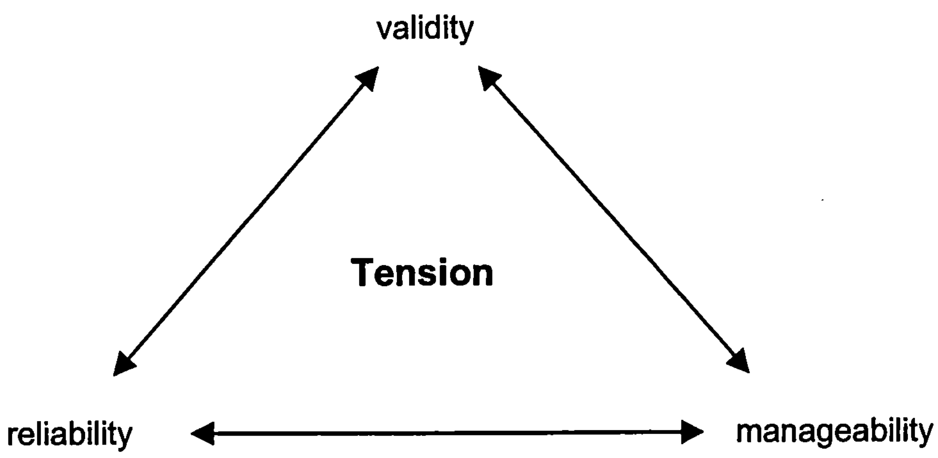


Figure 2.1: Tensions between the reliability, validity and manageability of assessment against US.

The process of quality assurance addresses these tensions through moderation procedures that aim to achieve high validity and optimum reliability within the constraints imposed by manageability.

2.2 What is meant by moderation?

Records of student achievement are of interest to teachers, students, parents, employers and tertiary institutions. Each of these stakeholders may use assessment results for different purposes. For these results to be useful and meaningful, fair comparisons must be able to be made of the achievement of the student relative to one or more of:

- the previous achievement of the student
- the achievement of other students in the same class
- the achievement of students taking the same subject in other classes in the same school

- the achievement of students in other schools
- specified local course objectives
- national examination prescriptions
- nationally stated standards.

Moderation is the process that enables the assessor judgements that underpin these comparisons to be made consistently.

The Ministry of Education booklet, *Policy to Practice* (1994b: 48) expands on this by defining moderation as the:

> ... process which ensures that assessments made by different people in different places or at different times are comparable. The process starts with agreed objectives and can include common assessment tasks, common scoring schemes, shared marking or grading and discussion between people carrying out the assessment.

The moderation process for assessment against US has to ensure that students from different schools who are assessed by different teachers using a wide variety of assessment tasks and contexts, at varying times during the school year, have achieved a comparable standard of competence when they are credited with the same US. This definition of moderation is demonstrated in Fig 2.2

A key word in these definitions is *comparability*. The first impression of this term is unproblematic. But what does this term really mean in relation to moderation? Sadler (1986) states that there are two easily distinguished definitions of the word *comparable* recognised by compilers of dictionaries. He discusses these at some length as they apply to the context of school-based assessment. The first definition of the word comparable is " able to be compared". This meaning derives directly from the etymology of the word. When using the word in this context, Sadler writes the word as *compare-able.*

> It is used when there is some question as to whether it is possible to compare things which are not identical (Sadler, 1986: 6).

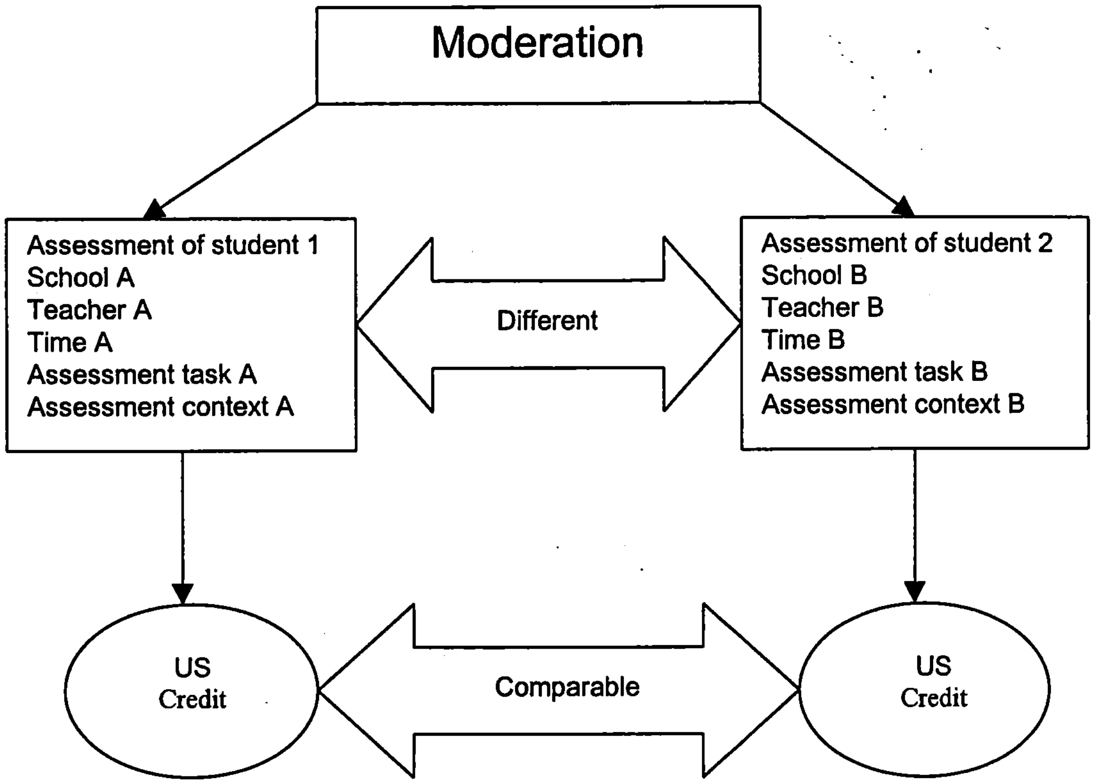


Fig 2.2: Achieving comparability through moderation

The second meaning of comparable is “more or less equal”. In this context Sadler writes the word as com-parable. He argues that *compare-ability* is a prerequisite to *com-parability*. Performances of students from different schools in a common examination can clearly be compared i.e. compare-able. The question of whether a cartoon drawn by a student is compare-able to a poster from a student at a different school, cannot be answered without resorting to the use of criteria. The criteria must be explicit and be independent of the context of the assessment. Sadler gives the example of comparing an egg and a glass of milk. An external criterion for comparison may be nutritional value or cholesterol content. Once this has been stated comparisons can be made. After the issue of compare-ability is settled, we can then ask whether the student performances are com-parable i.e. of the same standard. The implication of this argument for moderation of internal, SBA is that clearly stated context free external criteria are required. The performance criteria in US assessment are examples of such criteria.

Elsdon (1987: 17) argued that in addition to maintaining consistency between assessors and comparability between providers, moderation should have a "critical and developmental purpose". He views moderation as a crucial contribution to the professional development of the teachers involved in the moderation process. This has certainly been the case in US moderation. Moderators and teachers have been involved from the onset in writing parties, running courses for teachers, writing and checking the Nationally Prescribed Activities (NPAs) and providing feedback on the functioning of the moderation system. The interaction between teachers and moderators has been beneficial for the professional development of both teachers and moderators.

2.3 Why do we need moderation?

Why is moderation necessary? Why should teachers not use their own original assessments and raw assessment data without going through a moderation process? The answer to this question is related to the fact that teachers' individual assessment decisions depend on professional judgement. Sadler (1994: 31) states that professional judgement in turn depends on a teacher's understanding of:

(a) *the process of learning*: how children learn and develop

(b) *the process of making judgements*: how teachers come to make judgements

(c) *the outcomes of a judgement*: teacher insight into the outcome.

Since professional judgement depends on a large variety of factors it will differ from teacher to teacher.

> The purpose of moderating teacher assessments is to make the assessment fair by removing the effects of the teacher's judgement being different from that of his or her colleagues (Bennetts, Fairbrother and Willmott, 1986: 16).

Cowling (1994: 23) outlined the reasons moderation of the teacher judgements involved in assessment is necessary, under three major headings. These headings have been used as a framework for a more general discussion of the reasons for moderation.

- Wayward standards

The writer's experience as a national examiner and chief marker indicates that some markers are consistently lenient in their interpretation of an assessment schedule. Bennetts et al. (1986: 16) refer to this as "grading geese as swans". Other teachers are consistently harsh whereas still others are inconsistent and have an erratic interpretation of the assessment schedule which varies over time.

- Insufficient differentiation

Teacher assessments may not include sufficient coverage of the appropriate levels of the skills and content objectives of the curriculum or they may not provide students with sufficient opportunity to demonstrate that they have achieved the full range of required competencies.

- Variation in understanding of criteria

If teachers do not share the same understanding of the assessment criteria, variations in standards between assessors become evident. For instance, a teacher may not appreciate the difference between the higher level skill of "designing an investigation to solve a problem" and the lower level skill of "following instructions in setting up apparatus". Another example relevant to physics assessment is the skill of "recording and reporting". A teacher assessing this skill may interpret this to mean an emphasis on spelling and neatness whilst:

> … paying little attention to a candidate's attempt to relate results to scientific principles (Bennetts et al., 1986: 16)

2.4 Methods of moderation

There are three main methods of moderation that have traditionally been used for quality control of assessment. These are statistical moderation,

moderation by inspection and consensus moderation (Cowling, 1994 and Bennetts et al., 1986). In addition there are additional moderation techniques used for quality assurance of assessment. This section evaluates the suitability of each of these methods for the moderation of SBA.

2.4.1 Statistical moderation

This involves using a statistical method to scale the results of internal assessment to a distribution of marks arrived at using an external monitoring instrument. This method of moderation is used in New Zealand for the assessment of some subjects for senior secondary school qualifications. An example at the fifth form (Year 11) level is the use of a common reference test to establish comparability between schools for internally assessed SC subjects, such as, modular science, English or mathematics. The school's distribution of internally assessed school-based marks is scaled to the distribution of the performance of the same group of students in the reference test.

At the sixth form level, inter-school statistical moderation is carried out by allocating each school a pool of SFC grades which is based on the performance of the students in the previous year's SC examinations. In recent times this practice has become problematic. The abolition of scaling for SC, the growth in the number of internally assessed subjects at the fifth form level and new sixth form courses which do not have a fifth form equivalent, have made the allocation of grades based on the SC mark distribution increasingly difficult.

The prescription for the UB/S physics examination in Year 13 includes a 20% internally assessed component. This component is made up of 10% based on practical work carried out during the year and a special topic that contributes ten percent to the final mark. Teachers submit a school distribution of marks for the internally assessed component to the NZQA. This distribution of internally assessed marks has to reflect the distribution of marks that they consider their students will attain in the external examination. If the marks deviate too much from this distribution, NZQA scales the internal assessment marks to the school's distribution of marks in the external UB/S examination. Bennetts et al. (1986) refer to this as "the tolerance method". There are several problems associated with this method however. If schools spend little class time on the internally assessed component, and overemphasise the exam, students' exam marks will be inflated. This will, in turn, raise the internally assessed marks even though students have done little work towards this. The argument also works in reverse. If students spend too much time on practical work, their exam performance may suffer and consequently their internal assessment marks are reduced.

Schools which use ABA to assess the internal component, justifiably complain when their marks are scaled. These marks reflect student skills and abilities, which cannot be assessed in a written exam, and the distribution of internal results does not necessarily match that of the external exam. For this reason they should not be scaled to a mark distribution generated by an external exam nor aggregated with an examination mark for that matter. For schools with a small number of students, standard deviations are meaningless and statistical moderation is suspect. Sass and Wagner (1992) considered statistical moderation to be unsuitable as a moderation technique for SBA.

2.4.2 Moderation by inspection

Bennetts et al. (1987: 11) describe moderation by inspection as that carried out by an external moderator who inspects at least some of the coursework. This can be achieved by the moderator visiting schools or by work being sent to the moderator for re-marking. The benefit of this is that variability between schools can be substantially reduced if moderators are able to visit the schools frequently and are able to discuss the progress of each pupil in great detail. Basically the moderator establishes that the assessor judgements made by teachers are neither too harsh nor too lenient.

This method places a lot of importance on the consistency of the decisions made by the moderator. For this purpose the work of the moderator should be sampled by a chief moderator. Sass and Wagner (1992) prefer moderation by inspection to statistical moderation. In order to make this approach more manageable they suggest that a sampling approach could be used to moderate student work and that a proportion of the assessment for a course could be moderated in any one year. They sum up:

> Moderation by inspectors is complicated, time-consuming and relatively expensive compared to statistical moderation, but can lead to improved quality of teaching and learning and other opportunities for professional development (p 48).

2.4.3 Group or consensus panel moderation

This is a form of moderation which involves teachers discussing assessment activities and reaching a consensus about the standard of the activity, the nature of the student evidence required, and the assessment schedule before the activity is administered. It may also involve the use of agreement trials to verify the assessor judgements on marked student work. The purpose of an agreement trial is to determine the consistency of moderator decision making on set moderation criteria for a common assessment activity. Willmott, Bennetts and Fairbrother (1987: 11) state that the role of an agreement trial is to provide a forum in which a number of teachers in conjunction with a moderator, grade sample student course-work with the aim of discussing how their assessor judgements relate to those of their colleagues. The aim of the agreement trials that were run during the pilot phase of the Oxford Certificate of Educational Achievement (OCEA) in England was wider, and involved:

> ... bringing teachers together to discuss the assessment scheme, make judgements of students' work against the criteria, and to train teachers in the use of the criteria (OCEA, 1985).

Cowling (1994: 23) proposed a model of consensus moderation, which involves moderators working on panels to discuss issues, such as:

- whether students have been given sufficient opportunity to demonstrate their mastery of the competencies being assessed
- their understanding of the competencies and the evidence required for demonstration of mastery
- what constitutes sufficiency of evidence
- portfolios of student work
- national moderation support.

Bennetts et al. (1987: 14) feel that consensus moderation is more effective than statistical moderation or moderation by inspection.

2.4.4 Additional approaches to quality assurance

There are three additional approaches to moderation associated with quality assurance of assessment. Harlen (1994) listed them as, defining criteria for assessment, exemplification and approval of institutions and centres.

One way of improving consistency of assessment between different providers is to define criteria or standards which can be used by teachers to assess against. This form of moderation is useful to establish consistency in SBA.

In England, Wales and Northern Ireland, the General National Vocational Qualification (GNVQ) is awarded on the basis of competency-based assessment against national standards. (Hearn, 1996b). In order to establish consistency of assessment between over 2000 centres, the qualifications are broken down into units. Assessment for the units is against pre-defined standards, which are stated in terms of specific learning outcomes (elements) and associated performance criteria. The performance criteria are used to determine if the elements have been achieved. The Scottish system operated by the Scottish Vocational Education Council (SCOTVEC), is similar to the

GNVQ system. The units which have been achieved by an individual are recorded on a Record of Education and Training (RET), and can be aggregated into nationally recognised Vocational Qualifications, a National Certificate or National Diploma. The New Zealand Qualifications Framework has adopted the unit-based approach of the SCOTVEC and GNVQ and the PUS are an example of the criteria referred to by Harlen (1994).

To achieve reliability the learning outcomes expressed in the criteria or standards need to be unambiguous and highly specific. Higher order learning outcomes are often difficult to express in behavioural terms and are in danger of being underrepresented. To validly assess the wide range of learning outcomes of a programme, a large number of criteria is needed. This, in turn, poses a threat to reliability. The tensions between validity and reliability create the need for frequent revisions of the standards.

Exemplification involves providing assessors with examples of quality assessment activities and marking schedules. This illustrates the interpretation of standards of assessment, the writing of assessment activities which enable students to demonstrate mastery of the standard and the writing of judgement statements. The provision of examples of marked student work, accompanied by comments on particular features used in making the judgement is an additional aid to reaching consistency between assessors. Examples of this include the training guide *Standards-based Assessment for the National Qualifications Framework: Physics* (NZQA, 1996b) and the *Assessment Guide: Physics* (NZQA, 1996c) which provide examples of good and poor quality assessment activities and exemplars of student work. The annual Nationally Prescribed Tasks are additional examples of high quality assessment activities.

Harlen (1994: 22) describes the approval of institutions and centres as another approach to quality assurance:

> ... by which the body responsible for certain awards approves an institution or centre for training as one which can provide the course or training and can carry out the assessment related to these awards.

The approval process includes the inspection of provider policies related to assessment, vetting of staff and ensuring that the institution has the financial and physical resources to deliver the course and assess it. An example of this form of moderation is the 'quality auditing of centres' used by the SCOTVEC. In New Zealand the process of approval or quality auditing of centres is called accreditation. Educational providers must be accredited before they can offer courses leading to qualifications which may be registered on the NQF.

2.5 When should moderation occur?

Cowling (1994) classified two approaches to moderation which are commonly used: end loaded moderation and front loaded moderation. Much of the published material dealing specifically with moderation (for instance Kempa, 1986; Secondary Exams Council, 1986; Bennetts et al., 1986) focuses on ex post facto moderation which follows marking of internally assessed course work components and, in particular, on statistical moderation of examination results. This is "end loaded moderation" in which procedures to ensure consistency and comparability of assessor judgements are carried out after students have completed the assessment activity. This type of moderation utilises quality control and has traditionally been used in norm-referenced internal assessment.

The second approach is "front loaded moderation" in which moderation procedures are applied at the assessment activity and assessment schedule design stage. This is done in order to ensure greater consistency of assessor judgements and reduces the need for ex post facto moderation. This form of moderation is an example of a quality assurance procedure which is typically associated with SBA.

In the UK, there are three phases in the quality assurance and control procedures for the GNVQ; the preparatory phase, the concurrent phase and the review phase (Hearn, 1996b).

The preparatory phase

This phase includes the:

- accreditation of assessment centres
- provision of a MAP
- specification of national standards
- training of assessors.

These aspects were shared by the GNVQ and the NQF, but Hearn (1996b) states that in New Zealand, training was offered to all assessors whereas for the GNVQ it was patchy.

The concurrent phase

The concurrent phase for the GNVQ includes an average of two visits per year by an external verifier and multiple choice external tests for all subjects. Students are required to get a pass in these but they don't contribute towards grading standard assignments. The NQF differs because it includes moderation of assessment tasks and schedules in advance and the contact between the moderator and provider occurs throughout the year. The annual voluntary meetings of Local Moderators with their providers also enables contact and collegial support between providers.

The review phase

The external verifier for the GNVQ verifies the marking of student work late in the year. This serves a quality control function. The New Zealand method of sampling throughout the year has more of a quality assurance function.

2.6 New Zealand-based research into different forms of moderation

Elley (1988: 1) emphasised the need for research into moderation when he said that:

> .... in the movement towards a greater measure of internal assessment in the senior school, New Zealand educators have been restricted by a lack of research and experience in methods of maintaining comparability of standards between schools when grading student achievement.

In 1992 the NZQA commissioned the New Zealand Council for Educational Research (NZCER) to undertake a survey of New Zealand and overseas research into moderation. The NZQA report *Investigations into Moderation* (Sass and Wagner, 1992) states that overseas research into the moderation of criterion-referenced or SBA was limited. The report provides a comprehensive overview of the development of moderation strategies for the New Zealand context. The trend in this development mirrors the drift from external norm-referenced exams to internal assessment against standards and the accompanying shift from quality control of assessment to quality assurance of assessment.

Prior to the introduction of SBA, the traditional forms of moderation used in New Zealand schools focused on achieving inter-school comparability by using statistical techniques. Elley and Livingstone (1972) saw ranking students in a norm-referenced system as the preferred basis for maintaining consistency of assessment standards between schools. They stated:

> From time to time it has been suggested that the problem of consistency of standards might be overcome by developing absolute or content-referenced scales. In such schemes a pupil's grade depends not on how he compares with his school colleagues, but what he knows or can do. Unfortunately, few subjects have been found to lend themselves to such refined scales at secondary school level. The rank

> order of merit obtained from assessments of a wide range of skills and objectives is still the only practical way of assessing in complex attainment areas (p 50).

They proposed and developed nine methods of moderation designed for use with norm-referenced assessment. These were:

1. moderation of sixth form assessments using SC results
2. moderation using tests of scholastic aptitude
3. moderation using tests of general achievement
4. moderation using "omnibus tests" of achievement
5. moderation by means of item banks
6. supplementing external exams with internal assessment
7. school accrediting, similar to University Entrance with a back-up exam
8. exchange of scripts between schools, similar to university honours papers
9. moderation by inspection.

Methods 5, and 7-9 are different in that they do not rely on statistical moderation. For this reason they are potentially useful to moderation of SBA. The other methods are more useful for moderation of norm-referenced assessment.

Assessment for SFC has spearheaded development and change in moderation because being completely internally assessed, it raised questions about the effectiveness of inter-school moderation. The traditional method for ensuring comparability between schools was the statistical moderation suggested by Elley and Livingstone (1972). Sass and Wagner (1992: 45) stated that this was:

> … meant to be a temporary device to monitor internal assessment at the sixth form level, *not* as a permanent device to moderate the grades given.

There was considerable teacher dissatisfaction with this form of statistical moderation (Sass and Wagner, 1992). In 1986 the Committee of Inquiry into Curriculum, Assessment and Qualifications (CICAQ) recommended that alternative moderation procedures for sixth form subjects be investigated (Recommendation 19, p.18).

2.6.1 The moderation trials

Five alternative systems of moderation for SFC were trialled. The aim was to develop a suitable model for moderation which could be used to establish SFC as a stand alone qualification. The Qualifications and Assessment Division of the Department of Education contracted researchers to evaluate alternative systems of moderation. Table 2.1 lists the evaluators and the types of moderation investigated.

Table 2.1: Summary of the subjects, type of moderation trialled and evaluators contracted

Subject	Type of Moderation	Evaluators (year of publication)
French	Work samples and exemplars	Peddie (1990)
Geography	Teacher panels	Davidson and Steer (1989)
Physical Education	Visiting moderator	Dept of Education (1989)
English	Group and skills sampling	Gilmore (1991)
Practical Art	Teacher panels	Codd, McAlpine and Hansen (1990)
Home Economics	Visiting moderator	McKay and Peters (1989)

The trial for SFC English involved:

> ... test-based moderation using group and skill sampling techniques (Gilmore, 1991: 15).

This form of moderation is a variant on Elley's (1988) recommendations on the use of reference tests to moderate internally assessed sixth form subjects. Out of eight elements selected, a sample of four elements was tested each year. Random samples of students at participating schools were given moderating tests based on one of the four elements. Teachers showed support for the use of test-based moderation, but did not want the results to be used to allocate a fixed pool of SFC grades. They preferred the results to be used to adjust the school's distribution only if inconsistencies in the school's mark distribution became apparent by comparison with the moderating test results.

The use of ABA accompanied by exemplars of marked student work, was investigated as a form of moderation in the French trial (1988). School-based moderation involved the comparison of student work with the graded exemplars of student work provided. No other form of external moderation was used. Peddie (1990: 135) found that the:

> ... degree of marker reliability was not sufficiently high to warrant strong support for the system trialled.

The physical education and home economics trials used ABA backed up by visitation as a method of inter-school moderation. In the physical education trial, moderator visits were made during the school year, whereas the home economics trial, used postal sampling of student work and end-of-year moderator visits. The major findings of the investigations (Dept of Education, 1989 and McKay and Peters, 1989) were that the formative moderator visits in the physical education trial were more successful than the end of year moderator visit and postal sampling used in the home economics trial.

The geography and practical art moderation trials used a consensus panel approach to moderation. In Geography, exemplars of assessment activities, assessment schedules and student work were shared among all schools in the trial. The unique feature of this trial was the evaluation of assessment tasks and schedules by cluster groups of teachers before these were used by schools. The trial was evaluated by Davidson and Steer (1989). For each school, the grade level distribution for each element was compared with the SC mark distribution. The means of each distribution were correlated using the Spearman Rank Order Correlation Coefficient. The correlation indicated that the type of moderation used in the geography trial achieved a degree of national inter-school comparability that was similar to the previous year's SC.

The main findings of the practical art moderation trial (Codd, McAlpine and Hansen, 1990) were that attitudes of teachers to ABA and consensus panel moderation were highly positive and that most teachers considered the model to be reliable and felt it could work nationally.

Walker (1990) carried out an additional survey of 12 Christchurch schools to investigate the moderation of the practical skills components of SFC biology, chemistry and physics courses. His study looked at moderation procedures used in individual schools and did not pursue the question of inter-school moderation addressed by the other trials. Most schools used statistical moderation to scale the student marks attained in the practical component of the course to a distribution of marks gained from examinations or tests. He found an average correlation coefficient of 0.6-0.7 between practical work marks and examination marks and argued that this was not sufficiently high to justify this form of statistical moderation and recommended the use of assessment against criteria as an alternative.

The outcomes of the above trials established that there was strong support for ABA and compared the effectiveness of each type of moderation in the

New Zealand context. This contributed to the shaping of the moderation systems adopted for the NQF and the PUS in particular.

2.6.2 Recommendations based on the moderation trials

The evaluators of the above moderation trials wrote a combined statement, *A Policy manifesto for Sixth Form Assessment* (Codd, McAlpine, Hansen, Gilmore, Peddie, Peters, McKay and Crooks, in Gilmore, 1991) in which they recommended policy changes to sixth form assessment. These recommendations were grounded in their research findings. These recommendations include among others that:

- The policy of allocating SFC grades on the basis of SC results should be terminated and that stand alone moderation procedures need to be developed for each subject.
- ABA criteria and profile reporting procedures should be developed for all sixth form subjects.
- Forms of moderation may vary between subjects and for different components within a subject.
- Subject teachers should meet in clusters to develop assessment practices.
- Pre- and in-service training in assessment needs to be provided for teachers.

Crooks (1990: 34) expressed the caution that some ABA criteria were so vague that they offered "little advantage over a normative system of grading" and that:

> ... even the most precise criteria .... still require interpretation in the context of a specific assessment task.

He argued that:

> ... without some form of moderation, consistency of standards, locally or nationally was unlikely to be achieved.

Crook's concerns are similar to the findings of the Viviani (1990) review of the system of competency-based assessment used in Queensland. This system of assessment against five levels of achievement was established following a Review of School Based Assessment (ROSBA) in 1976. Viviani (1990: 50) concluded that:

> there was widespread concern about comparability of assessment among students and their parents, among teachers, at some levels in universities and in other groups.

In relation to moderation Crooks (1990: 34) concluded that any moderation system should include communication and consultation among teachers on a regional basis to achieve consistency and enable teachers to share assessment ideas and exercises. He also said that this form of moderation can:

> ... act as a general mechanism for staff development, in relation to teaching and curriculum matters as well as assessment principles and practice (Ibid.).

2.7 Moderation and the New Zealand NQF

Sass and Wagner (1992) were commissioned by NZQA to investigate moderation strategies for achieving consistency with US. Their work synthesises a lot of the research quoted earlier in this chapter. They recommended:

> the consensus panel approach as the minimum moderation required for all US in the Framework and across all levels. We believe with few exceptions, it should be the basis for moderation of quality standards for all provider courses (Ibid. 49).

They also said that in addition other techniques may be necessary to maintain comparability between providers, and parity of performance standards. They suggested the following strategies for enhancing comparability between providers:

- Clear statements of instructional objectives.

- Training in applying competency or grading criteria.
- Exemplar materials containing set examples of work with associated grading criteria.
- General training in SBA techniques including how to apply grading criteria.
- Standards for recognition of prior learning (RPL).
- Internal quality management procedures for providers.
- Systems for recording and crediting results (Ibid.: 58).

Hudson (1982) quoted in Willmott, Bennetts and Fairbrother (1987: 12) supported integrating a variety of moderation procedures. He argued that moderation should not be seen

> .... as a series of discrete activities, but as an integral whole carried out under the supervision of the Chief and Assistant Moderators. These processes start with the vetting of syllabuses and question papers, and the participation in agreement trials and conclude with the validation of internally applied grading standards by external assessors and moderators.

Having reviewed the New Zealand moderation trials and some of the relevant overseas research into moderation, it is now appropriate to look at moderation for the NQF to determine the extent to which it reflects and incorporates research findings and recommendations. The moderation system for the PUS was designed to incorporate appropriate strategies from this range of possible moderation techniques.

2.7.1 Moderation of the PUS

The Science MAP (Appendix 5) outlines the moderation process used for the Level 1-3 PUS and was designed by NZQA on behalf of the Science and Technology Advisory Group (STAG). It employs aspects of the consensus panel approach used in the geography ABA trials. This plan was first implemented in 1996 when the PUS were trialled and is administered by a network of Local, Regional and National Moderators.

The Science MAP (NZQA, 1995a: 5) describes moderation as a quality assurance process. It says that the aim of this process is:

> to ensure valid, fair and consistent assessment decisions.

The MAP consists of ten complementary components:

1. Accreditation of providers

Before a provider can offer the PUS, they have to be accredited by NZQA. The booklet *Standard-Based Assessment for the National Qualifications Framework: Physics* (day 2) states that the accreditation process requires that the provider has policies and procedures in place which cover:

- The development and evaluation of teaching programmes
- Financial and physical resources
- Staff selection, appraisal and development
- Student entry
- Student guidance and support
- Assessment of practical and work-based components
- Assessment and reporting. (NZQA, 1996b: 6)

2. A national network of moderators

The MAP is administered by a national network of Local and Regional Moderators under the oversight of a National Moderator. John Boereboom, a Senior Lecturer in physics at the Christchurch College of Education was appointed in 1995 as the first National Moderator. In addition four Regional and eleven Local Moderators were appointed. These numbers were increased in subsequent years as the number of providers offering courses that are assessed against the PUS grew. Each moderator received three days training.

Each Local Moderator is assigned a group of providers. The role definition for Local Moderators (NZQA, 1995c) outlines their key responsibilities and states that they will:

- work with NZQA to ensure consistent interpretation of the US, in consultation with providers and other moderators
- ensure that moderated assessment activities and assessment schedules prepared by providers within his/her local group of providers are consistent with national standards
- carry out review and verification of a sample of candidate work following the use of the moderated activities to ensure assessor judgements are consistent with the national standards
- conduct an annual voluntary meeting for providers in his/her group of providers and submit an annual report to the National Moderator.

In addition to carrying out all the tasks of the Local Moderator, the Regional Moderators check a sample of the moderation of assessment activities and schedules, and verification of student work carried out by the local moderators. They conduct an annual meeting for Local Moderators in their regional network and submit an annual report to the National Moderator. The Regional Moderators investigate disagreements between providers and Local Moderators and refer cases of disputes that cannot be resolved to the National Moderator (NZQA, 1995d).

The role of the National Moderator (NZQA, 1995e) is similar to that of the Regional Moderators but contains additional responsibilities. The National Moderator meets annually with the Regional Moderators and has the final responsibility for the interpretation of the US. These are published in a regular newsletter. The National Moderator collates the annual reports submitted by Local and Regional Moderators and submits an annual report on the performance of the moderation system to NZQA.

3. Ensuring consistent interpretation of the US

It is essential that assessors receive guidance on the interpretation of the PUS and the design of assessment activities. Providers that offer US are allocated a Local Moderator who monitors and samples their assessment against the PUS. The work of the Local Moderators is sampled by Regional Moderators who in turn have their moderation work sampled by the National Moderator. Any interpretation requests are channelled to the National Moderator who produces a regular newsletter for moderators that contains a record of these interpretations.

4. An item bank of assessment activities

The Assessment Guide: Physics (NZQA, 1996c) was produced by groups of physics teachers during two five-day writing parties as a resource for teachers and an aid to interpreting the US. It contains an item bank of assessment activities at levels 1, 2 and 3. These assessment activities can be used without modification or adapted to suit the assessment needs of a particular provider. The assessment activities help signal the standard and are to be used as a guideline for providers designing their own activities.

5. Moderation of assessment activities and schedules for 25% of US being assessed before the activities are administered

At the commencement of each year, each provider intending to assess against the PUS is required to submit an assessment plan, which outlines their assessment programme for the year, including unit numbers and approximate assessment dates. Each year the National Moderator selects two US that are the preferred US for moderation. If a provider does not offer either of these, they can negotiate alternative standards with their Local Moderator. The aim of this is that approximately 25% of a provider's assessment programme will be moderated each year. This sampling process ensures that over a four year period all of the US offered by a provider will be moderated.

Providers are required to submit all of the assessment activities they intend to use to assess students against the nominated PUS before they are administered. The Local Moderator has three options. The assessment activity can be approved for use, be approved subject to some recommended changes or it can be requested that the activity be resubmitted. This process aims to ensure that the assessment activities used accurately reflect the relevant US.

6. *Raising assessor expertise*

All physics teachers in the 1996 level 2 PUS trial were offered three days initial training in:

- the NQF
- principles of SBA
- writing of PUS assessment activities
- writing of assessment schedules
- recognising examples of good and bad assessment activities
- the moderation system
- implementation of the new system.

The interaction of providers with the moderation system also raises the expertise of assessors by giving them an opportunity to discuss their assessment activities with a colleague outside the school. This is especially beneficial when they are the only physics teacher in the school.

7. *Training of moderators*

In addition to the three days of assessor training the physics moderators received a further two-day training programme. The training of physics moderators was undertaken concurrently with the training of science, chemistry, biology, agricultural science and religious education moderators. This cluster of subjects initially came under the oversight of NZQA assessment and moderation officer (AMO) Kate Colbert. Initial training sessions were held in Auckland and Wellington in November 1995 followed by a further training session in Wellington in April 1996. A total of 20 moderators were trained in these three sessions. The training programme was designed by the AMO in conjunction with the National Moderators. The training programme for the two days included the following sessions:

- familiarisation with terminology and the PUS
- understanding the physics MAP
- practice moderation of assessment activities
- practice verification of assessor judgements
- administration
- interaction with provider issues
- dealing with providers' non-compliance
- unresolved issues.

8. *Verification of assessor judgements*

After an assessment activity which is being moderated has been administered to students, the assessor submits a sample of six marked student scripts to their Local Moderator. The sample contains three scripts of students who have met the standard and three scripts of students just below the credit boundary. The Local Moderator may recommend that extra credit be given for sections that have been marked to an inflated standard or signal a requirement to tighten up marking procedures for future assessments. Once credit has been awarded by the provider, it cannot be retrospectively revoked by their Local Moderator.

9. *Nationally prescribed activities (NPA) for US selected by the National Moderator and the NZQA.*

In addition to the verification of assessor judgements for locally designed and administered assessment activities there was also a NPA each year. The NPA is internally assessed according to a national assessment schedule and is a further check to ensure national comparability between providers. The NPA was abolished in 1998 in response to research carried out by Moderation Services division of NZQA which showed that teachers found the restricted time period in which the NPA could be administered inconvenient.

The research did however point out the need for having externally designed pre-moderated activities and schedules for a given US. In 1998 pre-moderated activities were launched for some subjects.

10. A postal moderation system

The Science MAP uses a postal system to sample assessment activities and schedules used by assessors.

2.8 The research questions

The introductory chapter of this thesis discussed how the paradigm shift, from a norm-referenced model of assessment to a broader model of SBA has led to a Janus type assessment policy and associated political uncertainty about the introduction of the Framework. The debate about the introduction of US in the senior secondary school highlighted a range of unresolved issues related to the quality management of assessment against US. Assessment against the Level 2 PUS was selected as a context to evaluate the extent and relevance of these unresolved issues. MacDonald (in McCormick and James, 1983: 172) defined educational evaluation as:

> ... the process of conceiving, obtaining and communicating information for the guidance of educational decision making with regard to a specified programme.

To provide a structure for this evaluation, the current chapter identified validity, reliability and manageability as traditional key quality indicators of assessment and described how these indicators needed to be reconceptualised for evaluating the quality assurance process for assessment against the Level 2 PUS. This provided the background for the formulation of the following meta-question on which the current research is focused:

Is assessment against the PUS a valid, reliable and manageable way of assessing the achievement objectives of Physics in the New Zealand curriculum?

This question was broken down into three main research questions, which in turn were broken down into specific subsidiary research questions that link the unresolved issues identified by the public debate to the quality indicators. The questions were stated in a way that enabled them to be directly addressed by the research.

1 Is assessment against the PUS a valid way of assessing the achievement objectives of Physics in the New Zealand curriculum?

This general question about validity was broken down into the following specific questions related to the curriculum fidelity (including content and construct validity), concurrent validity, validity of the reporting process and consequential validity of assessment against the PUS.

a) Do the Level 2 PUS enable teachers to assess the full range of skill and content objectives of Level 7 of Physics in the New Zealand Curriculum?

One of the criticisms directed at the PUS was that they do not enable teachers to assess all the skills and content achievement objectives of the New Zealand physics curriculum. A survey of physics teachers carried out by the Education sub-committee of the New Zealand Institute of Physics (1996), found that 48 % of the 192 physics teachers surveyed felt that assessment against US was appropriate for assessing practical skills. In contrast, 72 % felt that they were inappropriate for assessing higher order thinking skills.

b) Is the number of credits for each Level 2 PUS proportional to the class time allocated to preparing students for assessment against those standards?

One of the questions raised about the PUS is whether the number of credits assigned to each US is appropriate. It is important that the number of credits for each level 2 and 3 US is proportional to the instructional time in a year's programme spent on the corresponding content. If this is not the case, there is a danger that assessment demands can end up driving and distorting the physics course.

c) How does the distribution of Level 2 PUS credits compare with the distribution of SFC grades?

The aim of this question was to establish whether assessment against the level 2 PUS has concurrent validity with traditional assessment for SFC.

d) How valid is the process of reporting student achievement for assessment against the Level 2 PUS?

This question addresses the validity of the reporting process of assessment against the level 2 PUS. Sadler (1993: 12) commented that in Queensland:

> … no comprehensive, scientifically collected and suitably analysed data have been available to date to provide a basis for assessing just what level of confidence could reasonably be placed in the levels of achievement.

This is an area addressed by the current research.

e) What is the impact of assessment against the Level 2 PUS on student learning and teachers.

This question focuses on consequential validity by investigating the impact of assessment against the Level 2 PUS on student learning, student motivation, student enjoyment of Year 12 physics, teacher enthusiasm and classroom teaching.

2 Does the MAP associated with the PUS achieve an acceptable level of national consistency of assessment and comparability between providers?

This question addresses the reliability of assessment against the PUS. For the new system of assessment and certification to be nationally and internationally acceptable there needs to be a high level of public, teacher and student confidence that the moderation system is achieving comparability between providers. This, in turn, depends on the consistency between moderators. Some avenues for research into moderation have been identified from overseas experience.
Sadler (1993: 12) suggested the following three possible approaches to research into the comparability of criterion based assessment:

> empirical research into the robustness of the system of assigning levels of achievement to student work, using reference testing to estimate the consistency of standards across schools and researching the criteria and standards for assigning grades to student folios.

The following research questions were designed to investigate to what extent the various aspects of the moderation system are effective in contributing towards achieving comparability between providers. The questions are presented in the chronological order of their relationship to the operation of the MAP.

a) How effective was the moderator training in orienting moderators to the MAP?

The training of the moderators is the first step in establishing consistency. Moderators' opinion on the effectiveness of the training process is a useful indicator of the extent to which consistency can be established.

b) How satisfactory are each of the following aspects of moderation in achieving comparability between providers:

- ***internal moderation procedures***
- ***moderation of the assessment plan***
- ***moderation of assessment activities***
- ***verification of assessor judgements***
- ***check moderation***
- ***communication within the moderation system***
- ***moderator and provider meetings?***

These questions can be investigated by surveying teacher and moderator opinions on the extent to which various aspects of the MAP contributed to achieving comparability between providers.

c) What is the role of the NPA in the MAP?

Coupled with this main question are the subsidiary questions:

- ***How satisfactory is the NPA in achieving comparability between providers?***
- ***When during the year should the NPA be held?***
- ***What should the results of the NPA be used for?***
- ***Is the NPA necessary?***

d) What is the consistency of front-end local moderation of assessment activities?

The longitudinal trend in the consistency of moderator decisions can be investigated using annual moderation agreement trials. In the UK agreement trials have been used to investigate consistency (OCEA, 1985). These are applicable to the New Zealand context and have been incorporated in the design of the research questions and method of the present study.

e) What is the national approval ratio of assessment activities?

The proportion of activities that are submitted to moderators and approved immediately without the need for re-submission can be used as an indicator of consistency between assessors. This can be monitored longitudinally.

f) What is the consistency of end-point assessor judgements between different providers?

This question aims to investigate whether there are any longitudinal trends in the consistency of end-point assessor judgements between different providers.

g) What are the threats to achieving consistency through the MAP?

This question aims to identify the factors that are threats to achieving comparability between providers with a view to improving the moderation system.

3. Is assessment against the Level 2 PUS manageable?

This question was broken down into the following subsidiary questions.

a) Is the workload associated with administration, moderation and assessment of the Level 2 PUS manageable for teachers, students and moderators?

The US debate outlined in Chapter 1 and the PPTA salary claim in 1996, highlighted workload as one of the issues of concern to teachers. The NZIP (1996) survey found that 66 % of the 192 teachers surveyed felt that the workload associated with assessment against the PUS was unsustainable.

b) Are the resources and support provided by NZQA sufficient for the implementation of the Level 2 PUS?

Resources and support include such aspects as, the professional development of teachers, the quality and accessibility of the Level 2 PUS, the *Assessment Guide: Physics*, the NPA and pre-moderated activities and communication related to implementation.

c) What are the school and physics departments based issues related to the implementation of assessment against the Level 2 PUS?

These issues might differ from those facing the individual physics teacher and need to be investigated by detailed longitudinal case studies.

Chapter 3 outlines the research methods that were used to investigate these research questions and describes the subjects and longitudinal structure of the research.

Chapter 3

The Research Method

This chapter explains the research design, describes the subjects of the research and outlines the procedures used to select the samples. It provides a detailed description of the longitudinal structure of the research and the instruments used to investigate the research questions. The chapter concludes with an outline of the types of data gathered for the research and the statistical and qualitative methods used to analyse the results.

3.1 The research design

Since assessment against the PUS is in its infancy, it is not clear whether the issues raised by the public debate (§ 1.3) are:

> ... merely the teething troubles of adjusting to a new paradigm or because of fundamental weaknesses (Smithers 1997: 23).

For this reason, the research questions cannot be validly answered by taking a snapshot of teacher, student and moderator opinion in the first year of implementation but need to be investigated longitudinally. The research commenced with the initial trial of the Level 2 PUS in 1996 and concluded in 1998.

The most appropriate way to investigate the research questions was to survey moderators, teachers and students who had direct experience of assessment against the PUS. Each year, the subjects for the research included all of the Level 2 PUS moderators, national samples of Teachers in Charge (TIC) of physics and national samples of Year 12 physics students. In addition, longitudinal case studies were used to monitor the implementation of the Level 2 PUS at five Canterbury schools.

Qualitative and quantitative methods of data collection included annual moderator, student and teacher questionnaires and annual moderator and end-point assessor judgement agreement trials. Annual interviews were carried out with the TIC of Year 12 physics at five Canterbury schools. Table 3.1 contains an outline of the longitudinal structure of the research and the data collection instruments that were employed over the period of the study.

The research also contains an action research component. During the period of the research, the author was the National Moderator for the PUS and submitted annual reports to the NZQA that made research-based suggestions for improving the moderation system. In addition the author was contracted in in 1996 to review the PUS and used information gathered by the research to contribute to this process. The changes to the moderation system and PUS were monitored in subsequent years. This application of research findings to educational practice fits Burns' (1994: 293) definition of action research:

> Action research is the application of fact finding to practical problem solving ... with a view to improving the quality of action within it.

Action research also:

> ... involves direct co-operation between researchers and educational practitioners (Brock-Utne,1980)

This is a feature of the present research and provides:

> ... a means of improving the normally poor communication between the practising teacher and the researcher (Burns, 1994: 300).

Table 3.1: Outline of research design and instruments employed

	Subjects and research instruments		
Year	**Moderators**	**Teachers**	**Students**
1996	• Moderator questionnaire. • Consistency of moderation agreement trial: sample activity to determine consistency of moderation.	• Trial school teacher questionnaire. • Control school teacher questionnaire. • Interviews with the teacher in charge of physics at Canterbury schools in case study. • End-point assessor judgement agreement trial based on the 1996 NPA	• Trial school student questionnaire. • Control school student questionnaire. • Comparison of the 1996 PUS credit distribution with the 1996 SFC grades distribution.
1997	• Moderator questionnaire. • Consistency of moderation agreement trial: sample activity to determine consistency of moderation.	• Teacher questionnaire to all schools that assessed against the PUS and all schools that did not assess against the PUS • Interviews with the teacher in charge of physics at Canterbury schools in case study. • End-point assessor judgement agreement trial based on the 1997 NPA	• Student questionnaire for students assessed against the PUS and students assessed for SFC only. • Comparison of the 1997 PUS credit distribution with the 1997 SFC grades distribution.
1998	• Moderator questionnaire. • Consistency of moderation agreement trial: sample activity to determine consistency of moderation.	• Teacher questionnaire to all schools that assessed against the PUS and all schools that did not assess against the PUS • Interviews with the teacher in charge of physics at Canterbury schools in case study.	• Student questionnaire for students assessed against the PUS and students assessed for SFC only.

3.2 The subjects of the research

Table 3.2 contains an overall summary of the subjects, target samples and sample sizes for each year of the research. The nature of the subjects and the composition of the samples are discussed in detail in subsequent sections.

Table 3.2: National target sample sizes (N), actual sample sizes (n) and response rates to questionnaires (%) (1996-1998)

	1996			1997			1998		
Subjects	N	n	%	N	n	%	N	n	%
Level 2 Physics Moderators	17	14	82	27	25	93	24	22	92
Teachers in Charge of Year 12 Physics	70	52	74	430	159	37	430	150	35
Year 12 physics students	1400	848	61	1000	507	51	1000	526	53

In addition, longitudinal case studies were carried out at five Canterbury secondary schools.

3.2.1 The Level 2 PUS moderators

For each year of the research the target sample of moderators consisted of all of the Level 2 physics moderators who were contracted by the NZQA. The actual annual samples consisted of the moderators who responded to the questionnaires. Table 3.3 outlines the sample sizes and response rates of the Level 2 Physics moderators. N represents the total number of moderators working nationally at each tier of the moderation system each year, n represents the number of moderators who responded. The response rates are expressed as a percentage.

Table 3.3: The Level 2 Physics moderators

	The Level 2 physics moderators								
	1996			1997			1998		
	N	n	%	N	n	%	N	n	%
Regional Moderators	4	4	100	4	4	100	4	4	100
Local Moderators	13	10	77	23	21	91	20	18	90
Total	17	14	82	27	25	93	24	22	92

3.2.2 The Teachers in Charge of Year 12 physics

Each year of the research, the target population of teachers consisted of TIC of Year 12 physics at all New Zealand secondary, composite and area schools. They were targeted because they are responsible for the implementation of assessment policy and new developments in Year 12 physics assessment. They are acquainted with NZQA communications related to Year 12 assessment and the administrative requirements associated with assessment for SFC and the PUS. Each year the target population consisted of two main sub-sets, namely teachers that assessed against the Level 2 PUS and teachers that assessed for SFC only.

In 1996 an initial target sample of 35 TIC of Year 12 physics at schools that participated in the PUS trial and a matched control group target sample of 35 TIC of Year 12 physics at non-trial schools were selected. Stratified random sampling was used to select target samples that conformed to the target percentages established by the breakdown of the national population of all 1996 trial schools into the categories of type, administration, gender and size.

This procedure ensured that the sample of trial schools and the sample of non-trial schools were matched and of a similar composition to the national distribution of trial schools. The matched composition of the samples meant that valid comparisons could be made between them and that the findings of the research could be interpreted as representative of national trends.

The database provided by the Moderation Services division of the NZQA indicated that 122 schools had signed up to trial the PUS in 1996. Table 3.4 contains a breakdown of the national trial school population by type of administration, gender and size. The table headings represent the following groups of schools.

N_{trial} represents the total number of schools in each category of the national trial school population. The characteristics of these trial schools were obtained from the *Directory of New Zealand Schools and Tertiary Institutions* (Ministry of Education: 1995).

$\%_{target}$ represents the percentage of schools in each category of the trial school population.

n_{target} represents the target number of schools in each category of the trial and non-trial school samples. The target samples conform to the identical percentage composition of schools in each category established by the breakdown of the national population of trial schools.

$n_{\text{trial and non-trial}}$ represents the actual numbers of schools that responded in each category of the trial and non-trial school samples.

$\%_{\text{trial and non-trial}}$ represents the percentages of schools in each category of the trial and non-trial school samples of respondents.

A comparison of the percentage compositions of the trial school sample and $\%_{target}$ shows that they are closely matched. This indicates that the trial school sample is probably representative of the national population of trial schools. The close match between the percentage compositions of the trial and non-trial school samples enabled valid comparisons to be made between the two groups

Table 3.4: Target population of trial schools' characteristics, and trial and non-trial school sample composition (1996)

%	Category	National population of trial schools (N=122)		Target sample (N=35)	Sample of trial school respondents (N=26)		Sample of non-trial school respondents (N=25)	
		N_{trial}	$\%_{target}$	n_{target}	n_{trial}	$\%_{trial}$	$N_{non\text{-}trial}$	$\%_{non\text{-}trial}$
Type	Composite (Year 1-13) and area (Year 7-13)	11	9	3	2	8	2	8
	Secondary	111	91	32	24	92	23	92
Administration	Private	7	6	2	1	4	1	4
	Integrated	18	15	5	4	15	4	16
	State	97	79	28	21	81	20	80
Gender	Boys	20	16	6	4	15	3	12
	Girls	15	12	4	3	12	4	16
	Co-educational	87	72	25	19	73	18	72
Size	Small (< 501)	32	26	9	6	23	6	24
	Medium (501-850)	52	43	15	11	42	11	44
	Large (> 850)	38	31	11	9	35	8	32

In 1997 and 1998 the composition of schools assessing against the PUS changed with schools commencing or withdrawing from assessment against the PUS. Information on which form of assessment schools were using was difficult to obtain at the start of each year. To get full coverage, questionnaires were sent out to **all** New Zealand secondary, composite and area schools. The numbers of respondents were 159 teachers in 1997 and 150 teachers in 1998. This represents response rates of 37% and 34% respectively.

3.2.3 The Year 12 physics students

For each year of the study the population of students consisted of all Year 12 physics students in New Zealand.

In 1996, the target population of students consisted of all of the Year 12 physics students taught by the Teachers in Charge of Year 12 physics at each of the schools selected for the 1996 trial and non-trial school samples. Since class sizes were not known, each school in the sample was sent a bundle of 20 student questionnaires. Teachers were asked to copy or request additional questionnaires if needed. This represented a total mail out of 700 questionnaires for students at trial schools and 700 questionnaires for students at the matched sample of non-trial schools.

In 1997 and 1998 the samples of students were selected as follows. The total population of schools was divided into schools that had assessed against the Level 2 PUS and schools which had not assessed against the Level 2 PUS each year. Each of these groups of schools was listed alphabetically and a process of systematic sampling was used to select the schools to which student questionnaires were sent. Five hundred student questionnaires were sent to schools that assessed against the Level 2 PUS. A further 500 student questionnaires were distributed to schools that had not assessed against the Level 2 PUS. Since a number of schools requested additional questionnaires in 1996, the number of questionnaires distributed to these schools in 1997 was increased to 25. The remainder of schools were sent 20 questionnaires each. The covering letter contained the instruction that the questionnaires were to be administered to the Year 12 physics class taught by the TIC of physics.

3.2.4 The Canterbury schools for the longitudinal case studies

The purpose of the case studies was to follow up in more detail on issues raised by teachers in their responses to the annual teacher questionnaires and to focus specifically on school-based implementation and manageability issues related to assessment against the Level 2 PUS. The case studies were carried out over a three-year period commencing with the initial Level 2 PUS trial in 1996 to the final year of the study when the US had been more developed and established in schools. The sample of five Canterbury schools for the longitudinal case studies is a subset of the national sample of 1996 trial schools. The reason for selecting Canterbury schools was purely logistical and enabled the author who resides in Christchurch to conduct detailed interviews over a three-year period. The sample was selected to consist of a range of different types of schools and was composed of the following schools:

- School A was a large urban single sex boys' state secondary school of decile ranking 10. Each year of the study there were 4 physics teachers at the school and four Year 12 physics classes. Each of the physics teachers attended the three-day NZQA assessor training course in 1995. The teacher who was interviewed each year was the HOD science. The school participated in the Level 2 PUS trial in 1996 but discontinued assessing against the PUS part way through 1997.

- School B was a small urban co-educational state secondary school of decile ranking 4. Each year of the study there was one physics teacher at the school and one Year 12 physics class. The teacher attended the three-day NZQA assessor training course in 1995.The same teacher was interviewed each year. The school participated in the Level 2 PUS trial in 1996 and continued assessing against the PUS in 1997 and 1998.

- School C was a private single sex girls' school of estimated decile ranking 10. Each year of the study there were two physics teachers at the school and two Year 12 physics classes. Both physics teachers attended the three-day NZQA assessor training course in 1995.The teacher who was interviewed each year was the HOD science. The school participated in the Level 2 PUS trial in 1996, reduced their PUS assessment programme in 1997 and discontinued in 1998.

- School D was a small co-educational rural area school of decile ranking 7. Each year of the study there was one physics teachers at the school and one Year 12 physics class. The teacher attended the three-day NZQA assessor training course in 1995.The same teacher was interviewed each year. The school participated in the Level 2 PUS trial in 1996, continued with a full programme in 1997 but discontinued assessing against the PUS in 1998.

- School E was a large urban co-educational state secondary school of decile ranking 3. In 1996 there were two physics teachers at the school and two Year 12 physics classes. The TIC of Year 12 physics attended the three-day NZQA assessor training course in 1995 and was interviewed in 1996. In 1997 and 1998 there was only one physics teacher at the school and one Year 12 physics class. The school participated in the Level 2 PUS trial in 1996 but discontinued assessing against the PUS in 1997 and 1998.

The teachers who were interviewed at schools A to E are referred to in the text as teachers A to E.

3.3 The instruments

The instruments employed were specifically designed for the research and included annual:

- Moderator questionnaires.
- Moderator agreement trials.
- Year 12 physics teacher questionnaires.
- End-point assessor judgement agreement trials.
- Year 12 physics student questionnaires.
- Interviews with teachers in charge of physics.

These are discussed in greater detail in the following sections.

3.3.1 The moderator questionnaires

A copy of the 1996 moderator questionnaire is included in Appendix 6.The response rate was 82%. Similar questionnaires were administered in 1997 and 1998 to investigate the nature of longitudinal changes (Appendices 7 and 8). The response rates were 93% and 92% respectively. The 1997 and 1998 questionnaires contained additional questions to follow up on issues raised by moderators in previous years. Since the question numbers in the questionnaires changed from year to year as a result of the insertion of additional follow up questions, questions were tagged for quick identification in the text using a code. An example is the code 96MQ6 which refers to question number 6 in the 1996 moderator questionnaire.

3.3.2 The moderator agreement trials

At the commencement of each year of the study all of the Level 2 PUS moderators were asked to moderate a common sample assessment activity. The moderators carried out the usual process of moderation and recorded their moderation decisions on the standard NZQA pro-formas. The results were analysed to determine the degree of consistency of moderator judgements and enabled longitudinal comparisons to be made. Copies of the assessment activities used for the 1996, 1997 and 1998 moderation agreement trials are found in Appendices 9-11.

3.3.3 The Year 12 physics teacher questionnaires

In Term 4 of 1996 a questionnaire (Appendix 12) was administered to the sample of 35 TIC of Year 12 physics at schools which assessed against the Level 2 PUS. A separate teacher questionnaire (Appendix 13) was administered to a matched control sample of 35 TIC of Year 12 physics at schools that assessed for SFC alone. The return rate for each of the questionnaires was 74% and was adversely affected by the PPTA's Qualifications Framework freeze associated with the 1996 collective contract negotiations. It was further affected by the heavy workload associated with dual assessment for SFC and the PUS, and for some teachers, simultaneously implementing Science US and Mathematics US.

In 1997 and 1998 the same questionnaires were administered to all teachers in charge of Year 12 physics (Appendices 14 and 15). The questionnaires contained a separate section which was to be answered only by teachers who had assessed against the Level 2 PUS. A common core of questions was asked over the three years of the study to investigate the nature and extent of any longitudinal trends. Additional questions were included in 1997 and 1998 to investigate issues raised by teacher responses in previous years.

In 1997, teacher questionnaires were returned from 159 teachers. This included 102 teachers who did not assess against the Level 2 PUS and 57 teachers who did. Of the 57 teachers who assessed against the Level 2 PUS, approximately half (28) did not participate in the 1996 trial. The data relating to this subset has been included in the tables of results in brackets. Of the 102 teachers who had not assessed against the Level 2 PUS in 1997, 11 took part in the 1996 trial. Since this is only a modest percentage, the data were not reported separately. In 1998, 150 questionnaires were returned. This included 59 questionnaires from teachers who assessed against the Level 2 PUS and 91 who had not.

For each year of the study a sample of teachers who had not assessed against the PUS and a sample of teachers who had assessed against the PUS were asked to complete a checklist of the curriculum content they had covered with their classes. An analysis of these checklists was used to determine the extent of curriculum content coverage by schools assessing against the PUS and enable a comparison to be made with schools that did not assess against the PUS. In addition, physics teachers were asked if there were aspects of the course which they felt could not be assessed using PUS.

Since the question numbers in the questionnaires changed from year to year as a result of the insertion of additional follow-up questions, questions have been tagged for quick identification using the key described in Table 3.5.

Table 3.5: Key for the identification of questions in the teacher questionnaires

Code	Explanation
96TQT1	1996 teacher questionnaire for trial schools, question number 1
96TQN1	1996 teacher questionnaire for non-trial schools, question number 1
97TQ1	1997 teacher questionnaire, question number 1
98TQ1	1998 teacher questionnaire, question number 1

3.3.4 The end-point assessor judgement agreement trials

For each year of the study, an agreement trial was conducted to determine the level of consistency of end-point assessor judgements. In Term 3 of 1996 and 1997, a sample student answer for the 1996 and 1997 Nationally Prescribed Activity (NPA) was sent for marking to all of the schools that had used the activity in their assessment programme. Copies of the NPAs, assessment schedules and sample student answers may be found in Appendices 16 and 17 respectively. For each performance criterion in the supplied assessment schedule, the percentage of agreement between the end-point assessor judgements was determined. The mean percentage of agreement across all of the performance criteria assessed by the NPA was taken to represent the level of agreement of providers within the physics moderation system. Since agreement trials were carried out for each year of the research, it was possible to make longitudinal comparisons.

3.3.5 The Year 12 physics student questionnaires

In 1996 different student questionnaires were used for Year 12 physics students who attended the Level 2 trial and non-trial schools. Copies of these questionnaires may be found in Appendices 18 and 19, respectively. The questionnaires are identical except for the section of questions related to assessment against US which was omitted for the non-trial schools.

In 1997 the same questionnaire was used to survey Year 12 physics students at schools which had and had not assessed against the Level 2 PUS. A copy of the 1997 questionnaire may be found in Appendix 20. The questionnaire contained a separate section that was to be answered only by students who were assessed against US in 1997. The 1997 questionnaire was similar to the 1996 questionnaire but included a number of additional questions to follow up issues raised by the 1996 Year 12 cohort of physics students.

The 1998 questionnaire was similar to the 1997 questionnaire but included an additional section on students' knowledge about the framework. A copy of the 1998 questionnaire may be found in Appendix 21. Since the question numbers changed from year to year, they have been coded for quick identification in the text, using the key in Table 3.6.

Table 3.6: Key for the identification of questions in the student questionnaires

Code	Explanation
96SQT1	1996 student questionnaire for trial schools, question number 1.
96SQNT1	1996 student questionnaire for non- trial schools, question number 1.
97SQ1	1997 student questionnaire, question number 1.
98SQ1	1998 student questionnaire, question number 1.

3.3.6 Reliability and validity of the questionnaires

The questions and overall design of the moderator, teacher, and student questionnaires were pre-tested by asking three moderators, teachers and students respectively to complete, critique and comment on the questionnaires (Gay, 1996; McCormick and James, 1983). Small changes were made in response to this feedback to optimise the face and content validity of the questionnaires.

To encourage a high return rate, stamped addressed envelopes were provided. After the deadline for the return of the questionnaires follow up faxes were sent to encourage additional returns (Burns, 1994; Gay, 1996).

All the questionnaires used Likert scale closed questions combined with open ended questions designed to elicit a short response (McCormick and James, 1983). The validity of Likert scales used in the questionnaires depends on the honesty of the self reporting by the moderators, teachers and students. Burns (1994: 342) cautions that Likert scores:

> ... merely summarise the verbalised attitudes the subjects are willing to express.

This has limitations because it depends on what the respondents are willing to reveal about themselves. This was partly addressed by informing respondents that individuals would remain anonymous in the analysis and reporting of results. Since three groups of subjects were surveyed on similar issues, the triangulation of results provided information to evaluate the internal validity of the conclusions. The evaluation of the research findings in the context of the research literature provided additional support for the validity of the findings.

3.3.7 The interviews

The first round of interviews was conducted with the TIC of physics at the five Canterbury schools late in term 4 of 1996. The purpose of these interviews was to get a detailed overview of the issues related to the implementation of standards-based assessment in general and the Level 2 PUS trial in particular. Further interviews were conducted in term 4 in 1997 and 1998 in order to document the nature and extent of any longitudinal changes related to the implementation and evolution of this form of assessment in the case study schools. Best and Kahn (1998: 322) remarked that the validity of interviews is:

> ... greater when the interview is based on a carefully designed structure, thus ensuring that the significant information is elicited (content validity)

To achieve satisfactory content validity, a structured schedule of interview questions was prepared each year and discussed before use with a number of colleagues. Reliability was addressed by triangulating the interview data with the data obtained from the teacher questionnaires completed by the same teachers.

A copy of the questions used to conduct the annual interviews is contained in Appendix 22. Supplementary questions were asked where appropriate.

3.4 Data analysis procedures and tests of statistical significance

The processing of the data collected by the research depended on the properties of the data collected. This section outlines the three types of numerical data collected by the teacher, student and moderator questionnaires and explains the rationale for the selection of the statistical techniques used to analyse the data. In addition to the quantitative data, the research collected qualitative data in the form of written responses to the questionnaires and taped teacher interviews for the case studies.

3.4.1 Nominal scale data

Data in this category included subject classification type data, such as, type of school and gender. This type of data is nominal and was processed by frequency tabulation. To enable comparisons over time, the frequencies were converted to percentages.

3.4.2 Ordinal data

The questionnaires used Likert scales to survey student, teacher and moderator opinion. Burns (1994: 338) states that the advantage of this method is that it:

> produces more homogeneous scales and increases the probability that a unitary attitude is being measured, and therefore that validity (construct and concurrent) and reliability are reasonably high.

The scales have been constructed in such a way that the intervals on the scales are symmetrical on either side of a neutral mid-point. The intervals on the scale are not necessarily equal however.

The ordinal data gathered by Likert scale questions was presented as a frequency distribution of categorical responses. Since the number of respondents changed from year to year and question to question, frequencies have been converted to percentages for ease of comparison. Hansen (1996: 23) stated the following caution:

> Quite simply we have no way of knowing if what one teacher intends when they respond to a Likert scale is the same, or similar, or even quite different to what the next teacher intends his/her identical Likert ranking to be. When analysing data obtained from Likert scales it is useful to look at the skew of frequencies i.e. whether these are mainly clustered towards the higher or lower end of the scale.

This advice was followed in the analysis of the tables of results of Likert scale data.

Longitudinal trends in moderator responses were analysed using a Chi-square (χ^2) test of independence with the level for statistical significance set at 0.05. The null hypothesis in each instance was that there were no longitudinal trends from 1996-98. According to Burns (1994: 177) the χ^2 test is:

> ... most appropriate for the analysis of data that are classified as frequency of occurrence within categories.

He further stipulates that the categories must be mutually exclusive. The data collected met these assumptions. The χ^2 analysis was carried out on the actual frequencies of responses, not the percentage summaries. In some cases, categories of responses were combined to meet the requirement that the expected frequencies should be equal to or greater than five in at least 80% of the cells (Burns, 1994; 178). Footnote references indicate where this was done.

The teacher and student questionnaire data were broken down into various subsets. The teacher questionnaire data collected each year (1996-98), was broken down into responses from teachers who had assessed against the Level 2 PUS and teachers who had not assessed against the Level 2 PUS.

The analysis required comparisons to be made between the two groups of teachers, as well as, longitudinal comparisons.
The student responses were analysed to determine whether there were any differences in the responses from students who were assessed against the Level 2 PUS and SFC and those students who were assessed for SFC only. To investigate whether school size had an impact on validity and manageability, the student data were analysed to investigate differences in the responses related to school size (large, medium or small).

Chi-square analyses did not permit comparisons between subgroups of teachers and students. The required analyses could be achieved by using two factor Analyses of Variance (ANOVA's). This form of analysis is valid if the data meets the following requirements:

- the variables measured are normally distributed in the population
- the data represent interval or ratio scales of measurement
- subjects are selected independently for the research.

Gay (1996: 467) argued that:

> ... with the exception of independence, some violation of one or more of these assumptions usually does not make too much difference. In other words, the same decision is made concerning the statistical significance of the result.

Since the data obtained from Likert scales is ordinal it could be argued that it is inappropriate to use ANOVA tests of significance. However Burns (1994: 338) claims that:

> Many researchers, of course assume Likert scales provide interval data,

and that:

> We often resort to allotting numbers to all sorts of scales and assume they are equal interval scales, for example attitude scales, thereby allowing parametric procedures to be applied (Ibid: 132).

He also stated that parametric tests such as ANOVA's are relatively robust. This means that it is unlikely that the percentage probability will be inaccurate. For these reasons it was decided to use two factor ANOVA to investigate differences between means related to school size, type of assessment and longitudinal trends. A confidence level of $\alpha \leq 0.05$ was used for the ANOVA's. Whenever the ANOVA's showed a statistically significant difference, Scheffé post-hoc tests were used to make multiple comparisons between the means in order to determine exactly which means were significantly different. The confidence level was set at the $\alpha \leq 0.05$ level. The Scheffé procedure for post-hoc comparisons was selected because:

> It is very robust to violations of the assumptions typically associated with multiple comparison procedures

and was the:

> ... most conservative of the paired comparisons procedures available (Statview, 1994: 323),

on the statistical analysis computer package used.

Correlations between PUS credits and SFC grades for the concurrent validity investigation in Chapter 4 were calculated using the Spearman rank order method. A correlation coefficient was calculated for each school that practised dual assessment. All students who did not have a SFC grade ***and*** a credit total were left out of the Spearman rank order correlation coefficient calculations since the calculation of this statistic requires data arranged as matched pairs. The Learner Information Services division of the NZQA supplied the data for this analysis.

3.4.3 Interval data

Data, such as, the percentage agreement of end-point assessor judgements and the extent of curriculum coverage were interval data because these are measured in units which are of equal intervals. Those data were analysed using parametric techniques such as t-tests and ANOVA with a confidence level for significance set at $p \leq 0.05$.

3.4.4 Qualitative data

The qualitative data collected each year of the research consisted of written responses to questions in the moderator, teacher and student questionnaires, and the taped interviews with teachers at the case study schools.

Each year all of the written comments on the moderator questionnaires were analysed.

In 1996 all of the comments made by teachers were analysed. In 1997 and 1998 the large number of respondents (159-150) made it impractical to analyse the comments made on all questionnaires. Consequently the comments made on 50% of the questionnaires were analysed. This was achieved by arranging the questionnaires in alphabetical order and selecting every second questionnaire.

Because of the large number of student respondents each year (848-507-526), a sample of 10% of the students' questionnaires were selected for the analysis of written comments. These samples were formed by arranging the school sets of questionnaires in alphabetical order of school name and selecting every 10th questionnaire.

The quantitative data were used to indicate strength of feelings and opinions. The comments made by moderators, teachers and students were used to explore the range of opinions and to explain the trends in the quantitative findings. For each question, comments were collated under themes and issues. Representative comments were quoted to illustrate typical responses but frequencies of responses have not been included.

For the first round of interviews, summary notes were recorded during the interview by the interviewer. For the subsequent rounds, the interviews were taped and a verbatim transcript was made. The interview notes were then collated under themes and issues and presented in a more readable format. Verbatim quotes are inserted to illustrate particular points. Additional information used to write the case study reports was derived from the written answers to questions from the teacher questionnaires administered to the school as part of the national sample.

Chapter 4

Validity of Year 12 Physics Assessment

This chapter discusses the validity of Year 12 physics assessment by focusing on the research question:

> ***Is assessment against the Level 2 PUS a valid way of assessing the achievement objectives of Level 7 of Physics in the New Zealand Curriculum?***

This question was addressed by a longitudinal investigation into the curriculum fidelity, concurrent validity, validity of the reporting process and consequential validity of assessment against the Level 2 PUS. The investigation utilised two complementary sources of evidence. The first consisted of the opinions of teachers and students who experienced assessment against the Level 2 PUS. The second involved a comparison with SFC. SFC has been in operation since 1969. It is a longstanding award and teachers are experienced in assessing for it. Consequently it provides a useful benchmark for establishing the relative validity of the Level 2 PUS. It can be argued that assessment against the Level 2 PUS should be at least as valid as assessment for SFC. An investigation into the effect of school size on consequential validity and students' perception of the validity of reporting was incorporated in response to validity and manageability concerns raised in Chapter 2. The data for this chapter were obtained from annual (1996-1998) physics teacher and student questionnaires. National assessment data were provided by the NZQA.

4.1 Curriculum fidelity of the PUS

The investigation into the curriculum fidelity of assessment against the Level 2 PUS focused on the following subsidiary research question:

Do the Level 2 PUS enable teachers to assess the full range of skill and content objectives of Level 7 of Physics in the New Zealand Curriculum?

For assessment against the Level 2 PUS to have high curriculum fidelity, the Level 2 PUS must provide teachers with opportunities to assess all of the content outlined at Level 7 of *Physics in the New Zealand Curriculum* (Ministry of Education, 1994a). Furthermore it is essential that assessment against the Level 2 PUS does not impact negatively on curriculum coverage of teaching programmes and that the number of credits allocated to each PUS is in proportion to the classroom time and school programme weighting for the corresponding unit of work. In addition the Level 2 PUS must be valid for assessing practical work, problem solving and understanding of physics concepts appropriate to Level 7 of the physics curriculum. The following sections discuss these aspects of curriculum fidelity.

4.1.1 Curriculum representation

Since the Level 2 PUS are internally assessed it is difficult to get a direct indication of the content validity at a national level. However, curriculum representation is a measure of how well the Level 2 PUS reflect the curriculum. This provides a measure of the potential content validity of assessment against the Level 2 PUS. The curriculum representation of the PUS was investigated by surveying teachers and by comparing the curriculum coverage achieved by schools that had assessed against the Level 2 PUS with that achieved by schools that assessed for SFC only.

Each year teachers who had assessed against the PUS were asked how well they considered the Level 2 PUS reflected the content outlined at Level 7 of *Physics in the New Zealand Curriculum* (Ministry of Education, 1994a).[1] Table 4.1 contains a summary of their responses.

[1] 96TQT2a, 97TQ13, 98TQ12

Table 4.1: Teachers' opinions on how well the Level 2 PUS reflected Level 7 of *Physics in the New Zealand Curriculum* (1996-1998)

		Percentage of teachers		
		1996	1997	1998
5	Very well	22	20	22
4	Quite well	70	65	55
3	Not sure	8	10	15
2	Poorly	0	5	8
1	Very poorly	0	0	0
Number of responses		26	49	55

($\chi^2_{1996/1997/1998} = 3.83$, $df = 4$, $p > 0.05$)[2]

Table 4.1 illustrates that each year, a substantial majority (92%-85%-77%)[3] of the teachers surveyed felt that the Level 2 PUS reflected Level 7 of the physics curriculum either "Very well" or "Quite well". Over the same period, a small minority (8%-10%-15%) of teachers was "Not sure". No teachers felt they reflected the curriculum "Very poorly" and in 1997 and 1998 only a small proportion of teachers rated the curriculum reflection "Poorly".

The following typical comment made by a teacher in 1996 indicates why teachers felt that the Level 2 PUS reflected the curriculum well:

> The Physics Unit Standards cover all of the content areas of the curriculum and are a very good match with the achievement objectives. We rewrote our school scheme before the Physics Unit Standards were published and found that they mapped very well to our Year 12 physics programme.

The PUS were revised at the end of the 1996 trial year. The revision consisted of the adjustment of credit values, the deletion of the extended

[2] Categories 1-3 were combined to meet the requirement that the expected values should be greater than 5 in at least 80% of the cells (Burns, 1994: 178).
[3] Throughout this chapter percentages in table descriptions are listed in chronological order.

investigation PUS 6384[4] and a reduction of the number of performance criteria and range statements. Teachers' comments in subsequent years referred specifically to some of these changes.

In 1997 and 1998 the main reason given by the small minority of teachers that rated the curriculum reflection of the PUS "Poorly", was disagreement with the removal of the 'extended investigation' PUS. One teacher commented:

> The open-ended extended investigation was a key component of the new curriculum and an important part of our physics course. It should not have been deleted from the Physics Unit Standards. The generic science investigation US that replaced it is too general for a physics course.

Other comments questioned the validity of the inclusion of content drawn from Level 6 of *Physics in the New Zealand Curriculum* (Ministry of Education, 1994 a) in the Level 2 PUS. This content was included because a majority of students take a Year 11 science programme based on Level 6 of *Science in the New Zealand Curriculum* (Ministry of Education, 1993) and therefore miss out on a proportion of the Level 6 physics content. Consequently, Year 12 teachers who adhered strictly to Level 7 of the physics curriculum in designing their programmes, may not have adequately prepared students for assessment against the Level 2 PUS and may have felt that the content validity of assessment against the Level 2 PUS was less than satisfactory. This line of reasoning is reflected by the following comment:

> It is unfair that the Level 2 standards include material from the Year 11 course that I don't usually teach in Year 12 and which is not part of the curriculum at that level.

The Level 2 PUS are highly prescriptive in terms of content. Further evidence for curriculum fidelity may be found in the extent to which schools that assessed against the Level 2 PUS had covered the achievement objectives of Level 7 of the physics curriculum in their Year 12 physics courses each year.

[4] *Investigate a physical system to determine a relationship with supervision.*

Table 4.2 compares the mean percentage of Level 7 curriculum content coverage for schools that assessed against the PUS, and schools that assessed for SFC only. The percentage of curriculum coverage was calculated by asking teachers to complete a detailed and itemised curriculum topic checklist and expressing the number of topics covered as a percentage of the total number of curriculum topics listed at Level 7 of the curriculum.

Table 4.2: Comparison of the mean and standard deviation percentage curriculum content coverage between schools assessing against the PUS and schools assessing for SFC only (1996-1998)

	Percentage of curriculum coverage					
	1996		1997		1998	
	PUS	SFC	PUS	SFC	PUS	SFC
Number of schools	25	26	57	99	57	92
Mean	93	86	87	91	92	93
Standard deviation	7	14	21	16	15	13

Table 4.2 indicates that each year both SFC and PUS schools achieved a high level (approximately 90%) of curriculum coverage in their programmes. A single factor ANOVA showed that there was no statistically significant difference in curriculum coverage between schools that assessed against the PUS and schools that assessed for SFC only *($F(1,354)_{PUS/SFC} = 0.46$, $p > 0.05$)*.

Since the PUS are based on the new curriculum, schools that assessed against the Level 2 PUS in 1996, had to simultaneously implement the new curriculum and the PUS. Schools that did not assess against the PUS could delay implementation till 1997, the official year for implementation. The new curriculum introduced only minor content changes and any implementation delays did not impact significantly on curriculum coverage. There were some newly introduced topics such as, "the voltage divider", that were most frequently not covered by both groups of teachers because they were "unsure of what was expected".

4.1.2 Validity of the Level 2 credit allocation

Another criterion for judging the curriculum fidelity of the Level 2 PUS is the accuracy of the number of credits assigned to each PUS. This is likely to have an impact on the amount of class time teachers will allocate to preparing students for that PUS. If the credit allocations are inaccurate, the assessment demands could distort the year's programme and assessment rather than the curriculum could end up driving the teaching programme.

The following research question addressed the validity of the credit allocations for the level 2 PUS:

> ***Is the number of credits for each Level 2 PUS proportional to the class time allocated to preparing students for assessment against those standards?***

The most appropriate time to investigate this question was in 1996 because all the schools that participated in the trial were contractually obligated to offer a full programme of Level 2 PUS. In subsequent years there was no contractual obligation and consequently the number of PUS offered varied widely from school to school. Therefore the appropriateness of the credit allocation was investigated only in 1996 and not in subsequent years.

Teachers who had assessed against the Level 2 PUS were asked to record the number of class periods spent on preparing students for assessment against each of the Level 2 PUS. This was expressed as a percentage of the total teaching time in the year's programme and compared with the percentage that the number of credits for each PUS contributed to the total number of credits in the full year's programme. Only 18 of the 26 respondents completed this part of the questionnaire. The reason for this low response rate may be that the completion of this aspect of the questionnaire was time consuming and required detailed records.

Table 4.3 contains a breakdown of the mean number of class periods spent on teaching, assessing and re-assessing against each Level 2 PUS during the 1996 trial. The class periods were 50-60 minutes long. In column A, the number of credits allocated to each PUS is expressed as a percentage of the total number of credits for the year. In column B, the number of class periods allocated to each PUS is expressed as a percentage of the total number of class periods for the year. B-A represents the difference between columns A and B.

In analysing the table, differences of 3% or less were deemed to be acceptable and within the range of uncertainty of measurement. This cut-off was decided on after consultation with a number of experienced teachers and moderators who regarded this level of variation to be acceptable. Teaching time is generally allocated to achieve curriculum coverage and meet students' needs. Where a mismatch occurs, an adjustment in the number of credits rather than a change in teaching time allocation should rectify this. Accordingly, a mismatch of more than 3% was taken to mean that the number of credits allocated to the PUS was inappropriate and out of proportion with the percentage of the year's programme allocated to teaching, assessing and reassessing for that PUS.

Table 4.3: Comparison of the percentage of total credits and the percentage of class time spent teaching, assessing and reassessing for each Level 2 PUS (1996)

PUS	Credits	Percentage of total credits (A)	Mean number of periods spent teaching and assessing	Percentage of total time (B)	Percentage difference (B-A)
6378	4	11.4	22.2	14.5	3.1
6379	4	11.4	17.8	11.7	0.3
6380	3	8.5	16.5	10.8	2.3
6381	6	17.0	26.7	17.5	0.5
6382	4	11.4	19.3	12.6	1.2
6383	2	5.7	9.5	6.2	0.5
6384	4	11.4	11.4	7.5	-3.9
6385	3	8.5	9.0	5.9	-2.6
6386	3	8.5	10.3	6.7	-1.8
6387	2	5.7	10	6.5	0.8
Total	35	100	152.7	100	1.7

Table 4.3 illustrates that the mean percentage difference between time allocation and assigned credit value was only 1.7 %. The conclusion that may be drawn from this is that overall the credit allocation was reasonably accurate. The PUS credit allocations identified as inappropriate were PUS 6378[5] for which the number of credits was insufficient and PUS 6384[6] which needed to have a decreased number of credits allocated to it.

To provide further evidence to the above statistical analysis an additional more subjective approach was used. Teachers were asked which PUS in their

[5] *Demonstrate knowledge of motion in one and two dimensions.*

opinion had an inappropriate number of credits allocated to them.[7] Seven out of the 26 respondents agreed that PUS 6378, which included projectile motion, circular motion and translational motion contained too much content for four credits, but did not comment that the credit value for PUS 6384 was inappropriate.

Five respondents felt that although the credit value for PUS 6381[8] was appropriate, it would be more convenient to assess if it were split into two smaller PUS with the number of credits assigned to each adjusted proportionately. The reason given was that magnetism and electricity were taught as separate topics and that it was more practical to assess them separately. This can be achieved however without splitting the standard by simply assessing the separate elements related to electricity and magnetism on different dates.

4.1.3 Skills objectives

In addition to enabling comprehensive content coverage it is essential that the Level 2 PUS permit teachers to adequately assess the full range of skill objectives of a Year 12 physics programme. In 1997 and 1998, teachers who had assessed against the PUS were asked whether the PUS enabled them to assess the following key skills:

a) practical work skills[9]

b) the ability to solve physics problems[10]

c) the ability to explain physics concepts.[11]

---

[6] *Investigate a physical system to determine a relationship with supervision.*

[7] 96TQT4c

[8] *Describe, construct and determine unknowns for electrical and electromagnetic systems.*

[9] 97TQ4a, 98TQ4a

[10] 97TQ4b, 98TQ4b

[11] 97TQ4c, 98TQ4c, 96TQT4

a) Practical skills

For the Level 2 PUS to have high curriculum fidelity, they must include opportunities for teachers to assess the full range of practical work skills recommended at Level 7 of *Physics in the New Zealand Curriculum* (Ministry of Education, 1994a). Table 4.4 summarises teachers' opinions on whether the PUS met this criterion for curriculum fidelity.

Table 4.4: Teachers' level of agreement with the appropriateness of the Level 2 PUS for assessing practical skills in physics (1997-1998)

		Percentage of teachers	
		1997	1998
5	Strongly agree	23	17
4	Agree	46	51
3	Not sure	18	22
2	Disagree	11	3
1	Strongly disagree	2	7
No of respondents		57	59

($\chi^2_{1997/1998} = 0.98$, $df = 3$, $p > 0.05$)[12]

Table 4.4 shows that in 1997 and 1998 approximately two-thirds (69%-68%) of teachers surveyed "Agreed" or "Strongly agreed" that the Level 2 PUS were appropriate for assessing practical skills and about a fifth (18%-22%) were "Not sure". Only a small minority (13%-10%) felt that the PUS were inappropriate for assessing practical work. There were no statistically significant changes in this pattern of responses between 1997 and 1998.

The following comment made by a teacher in 1997 indicates why teachers considered the PUS to be appropriate for assessing practical work:

[12] Categories 1and 2 were combined.

> The standards enabled me to assess practical work skills that could not be assessed in a written examination. You can observe and tick off each skill progressively and students get a diagnostic record of what skills they have yet to master.

One of the problems identified by teachers was the time, energy and equipment required to assess the practical work skills of a large Year 12 physics cohort. In 1998, one teacher explained:

> Assessment of practical work is very time consuming. With 110 pupils in Year 12 physics this year, it is difficult to assess a large number of students at the same time. We don't have enough apparatus. Neither do we have enough teachers for students to be observed individually.

Experimental work in physics is often carried out in small groups of three or four students. Under these circumstances teachers found it difficult to ensure the authenticity of student work. One teacher explained:

> Assessment of experiments which require more than one student to carry out may not be valid unless the students collect the data in groups and analyse it individually under formal test conditions. Otherwise it is impossible to identify what work the students did themselves and what they copied from their partners.

To ensure the validity of assessment of practical work carried out in small groups, some schools required students to sign declarations stating that the work submitted for marking was completed without assistance.

A minority of teachers felt that while the PUS were suitable for assessing the observable practical skills, they were:

> … not valid for assessing higher level practical skills that involved analysis and synthesis of physical ideas.

This concern is examined in detail in Part c of this section.

b) The ability to solve physics problems

The ability to solve physics problems is a key learning outcome of the Year 12 physics course. Table 4.5 contains a summary of teachers' opinions on whether the PUS were suitable for assessing this learning outcome.

Table 4.5: Teachers' level of agreement with the appropriateness of the Level 2 PUS for assessing students' ability to solve physics problems (1997-1998)

		Percentage of teachers	
		1997	1998
5	Strongly agree	7	17
4	Agree	43	43
3	Not sure	14	9
2	Disagree	29	28
1	Strongly disagree	7	3
No of respondents		56	60

($\chi^2_{1997/1998} = 3.90$, $df = 4$, $p > 0.05$)

Table 4.5 indicates that in 1997 and 1998 at least half (50%-60%) of the teachers surveyed felt that the PUS were appropriate for assessing problem solving skills but approximately a third (36%-31%) felt that they were inappropriate for this purpose. Only a small minority (14%-9%) of teachers was "Not sure". There was no statistically significant difference in the pattern of teachers' responses between 1997 and 1998.

A typical reason given by teachers for considering the PUS to be appropriate for assessing problem solving skills is illustrated by the following response to the 1998 questionnaire:

> The Physics Unit Standards allowed me to assess all the types of physics problems that I usually assess for Sixth Form Certificate. The performance criteria guide students through a solution, encourage

> them to set their work out properly and have improved students' attention to detail. It also made me more conscious of my own white board work because students use it as a model.

The following comment made in 1997 was a typical response of teachers who felt that the PUS were inappropriate for assessing problem solving:

> The narrow focus of the Physics Unit Standards and the restrictive nature of the elements tend to steer me away from setting physics problems that assess a variety of skills. Problem solving is too multi-dimensional to assess using the Physics Unit Standards. They are appropriate for problems that require only a single step solution but less so for higher order problems that require a synthesis of ideas.

In 1997, some teachers felt that while the PUS helped students to focus on the steps involved in problem solving, the pedantic requirements of the assessment schedules sometimes prevented them unnecessarily from achieving credit and stifled creativity by failing to acknowledge novel and unconventional ways of problem solving. One teacher explained:

> Students become more focused on what they need to do, and must be able to do it before they can get credit. However, many students can solve problems successfully but are unable to maintain the nitpickingly, politically correct terminology which some problems and written answers require. Are units and significant figures parts of the process of problem solving? What is required is too rigid, it seems just as much about preparing the answer in the right format as opposed to actually solving the problem.

Teachers were concerned that the performance criteria emphasised minor presentation details at the expense of higher level problem solving skills. This would impact negatively on curriculum fidelity.

c) The ability to explain physics concepts.

Achievement objective 7.1 of *Physics in the New Zealand Curriculum* (Ministry of Education, 1994a) states that students should be able to:

Apply concepts and principles to explain physical phenomena.

A necessary condition for curriculum fidelity is that the Level 2 PUS enable teachers to assess this curriculum objective. Table 4.6 contains a summary of teachers' opinions on whether the PUS were suitable for assessing this curriculum objective.

Table 4.6: Teachers' level of agreement with the appropriateness of the Level 2 PUS for assessing students' ability to explain physics concepts (1997-1998)

		Percentage of teachers	
		1997	1998
5	Strongly agree	5	12
4	Agree	21	37
3	Not sure	32	20
2	Disagree	33	24
1	Strongly disagree	9	7
No of respondents		56	60

($\chi^2_{1997/1998} = 5.98$, $df = 4$, $p > 0.05$)

Table 4.6 shows that teachers' opinion on whether the PUS were appropriate for assessing students' ability to explain physics concepts was divided. In 1997 and 1998, a sizeable minority (26%-49%) of teachers surveyed felt that the Level 2 PUS were appropriate for assessing students' ability to explain physics concepts but a similar sized group (42%-31%) deemed them inappropriate for this purpose. Approximately a fifth to a third (20%-32%) were "Unsure". There were no statistically significant differences across the years. This pattern of responses indicates that there are doubts about the suitability of the PUS for assessing students' ability to explain physics concepts.

A possible reason for this uncertainty was reflected in comments made by teachers in the 1997 questionnaire. These indicated that:

> The highly specific nature of the elements of the Physics Unit Standards made them unsuitable for assessing higher order skills, such as, explaining concepts and ideas and divergent thinking.

Teachers remarked that a simple pass/fail grade was inadequate for reporting the outcome of assessment of a complex skill like 'explaining physics concepts'. The following typical comment made by a teacher in 1998 explains this in greater detail.

> It is very difficult to write judgement statements that cover all possible ways of explaining an idea. They provide little opportunity to acknowledge the wide possible range of correct ideas students may generate or the degree of correctness. The degree of correctness of an explanation is probably worthy of marks rather than just applying a pass/fail standard. Often students will be able to explain some parts of an idea in their own words that don't quite match the rigid requirements of the standard. I feel this is worthy of some credit rather than asking them to do it again.

This section discussed the curriculum fidelity of assessment against the Level 2 PUS by presenting a number of complementary research findings. These findings need to be synthesised to judge whether the curriculum fidelity of the Level 2 PUS is acceptable.

The research found that each year a substantial majority of teachers felt that the Level 2 PUS covered all of the content areas outlined at Level 7 of *Physics in the New Zealand Curriculum* well. A related finding was that schools that assessed against the Level 2 PUS achieved a high level (91%-93%) of curriculum coverage in their courses that was similar to schools that assessed for SFC only. In addition the Level 2 PUS credit allocation was found to be generally appropriate and in proportion to teaching time.

Each year approximately two-thirds of teachers agreed that the PUS were appropriate for assessing practical skills and at least 50% agreed that they were appropriate for solving physics problems but teachers felt restricted in assessing higher order problem solving skills. Overall teachers were uncertain about the suitability of the PUS for assessing the ability to explain physics concepts. However in 1998 almost half (49%) of teachers surveyed felt that the PUS were in fact suitable for this purpose. This may reflect increasing teacher experience in designing assessments that integrate several elements (learning outcomes) into a single question.

Based on the results presented in this section it can be concluded that assessment against the Level 2 PUS had generally satisfactory curriculum fidelity but doubts remain about their suitability for assessing students' ability to explain physics concepts and higher order problem solving skills. The rigid requirements of the performance criteria tended to render them less valid for this purpose and teachers felt that they stifled teacher creativity by over emphasising lower level presentation skills.

4.2 Concurrent validity of PUS credits

Since assessment against the PUS is a potential replacement for the long established SFC, it is important that it has high concurrent validity with assessment for SFC. This aspect of validity was addressed by the research question:

> ***How does the national distribution of Level 2 PUS credits compare with the distribution of SFC grades?***

This question was investigated by comparing the distributions of the total number of Level 2 PUS credits registered per student and the corresponding SFC grades at schools that practised dual assessment.

The Spearman rank order method of calculating the correlation coefficients was employed because SFC grades are not equal interval data. The expected correlation values are negative because the magnitude of the two scales increases in opposite directions. SFC grades are awarded on a scale that ranges from one to nine with one being the highest grade. Consequently the magnitude of the grades decreased with increasing attainment. In contrast the number of Level 2 PUS credits registered on the NQF for individual students increased with increasing attainment.

In the 1996 questionnaire some teachers referred to:

> ... a possible difference in the performance of boys and girls because girls are more fastidious with some aspects that the performance criteria of the Physics Unit Standards focus on, such as, presentation details like rounding, showing all working and stating answers in sentence form.

They surmised that:

> Boys miss out on credit more frequently than girls because of small mistakes and less attention to detail. These boys often score highly on SFC assessments but miss out on an element for trivial reasons related to carelessness of presentation.

These comments were investigated in 1997, by calculating separate correlation coefficients for boys and girls.

Table 4.7 shows a summary of the 1996 and 1997 correlations between the total number of Level 2 PUS credits registered on the NQF and the corresponding SFC physics grade for each school that practised dual assessment. The number of schools that registered both Level 2 PUS credits and a SFC grade on the NZQA database was 50 in 1996 and 65 in 1997. The 1998 data was not available at the time of writing.

Table 4.7: Correlations between the total number of PUS credits registered per student and the corresponding SFC grades for students at schools that practised dual assessment in 1996 and 1997

	1996	1997		
	All students	All students	Female students	Male students
Range of correlation coefficients (ρ_{max}-ρ_{min})	-1 to +0.22	-1 to +0.13	-1 to 0	-1 to +1
Median correlation coefficient ρ	-0.72	-0.75	-0.77	-0.72
Percentage of schools for which ρ's were negative and significant @ the 0.05 level	79%	80%	90%	80%

Table 4.7 shows that in 1996 and 1997 the median correlation between the SFC grades and the total number of Level 2 PUS credits attained by students at schools that practised dual assessment was high (-0.72 to -0.75). Each year approximately 80% of the school correlations were statistically significant at the 0.05 level. In 1997 the median correlations for boys, girls and students in general were similar and no further statistical testing was carried out to investigate gender effects.

Overall it may be concluded that assessment against the Level 2 PUS has high concurrent validity with assessment for SFC.

4.3 Validity of reporting student achievement in Year 12 physics

In addition to having satisfactory curriculum fidelity and high concurrent validity, it is necessary to establish whether the reporting process is valid. This was addressed by the following research question:

How valid is the process of reporting student achievement for assessment against the Level 2 PUS?

After consultation with experienced physics teachers and moderators the following criteria for judging the validity of the reporting process were formulated:

- PUS credits must be valid for indicating student relative to the achievement objectives of Level 7 of the curriculum.
- The process of awarding credit must be transparent and easily understood by students.
- The reporting process must be useful for describing student achievement and must recognise different levels of achievement excellence.
- The process must be diagnostic and provide guidance for students and teachers on how achievement can be improved.

These criteria were judged using two complementary sources of evidence. The first involved surveying teachers and students who had experienced assessment against the Level 2 PUS and were therefore in a position to give an informed opinion. The second involved a comparison of teachers' and students' opinions on the validity of the reporting processes for the PUS and SFC. Since the Level 2 PUS are a potential replacement for SFC it can be argued that a valid reporting process using PUS should be rated at least as valid as that for SFC. The following sections discuss the research findings related to each of these criteria for validity of reporting. Teacher responses on the 1996 questionnaire indicated concern about the statistical validity of the allocation of SFC grades for small classes and raised issues related to the manageability of assessment against the Level 2 PUS. For this reason, the effect of school size on the validity of the reporting process was also investigated.

4.3.1 Validity of PUS credits

Each year of the study, the validity of Level 2 PUS credits for describing student achievement in Year 12 physics was investigated by surveying teachers who had assessed against the PUS.[13] To provide a basis for comparison, each year teachers were asked a parallel question about the validity of SFC grades.[14] Table 4.8 compares teachers' views on the validity of these two systems for reporting achievement. The responses for SFC are broken down into responses from teachers who assessed for SFC only and teachers who practised dual assessment for SFC and the PUS.

Table 4.8: Comparison of teachers' opinions on the degree of validity of Level 2 PUS credits and SFC grades for indicating student achievement in Year 12 physics (1996-1998)

		Percentage of teachers								
		PUS credits			SFC grades					
					SFC only			Dual		
		1996	1997	1998	1996	1997	1998	1996	1997	1998
5	Very valid	9	9	9	8	14	14	4	7	14
4	Valid	43	41	49	62	63	54	57	43	44
3	Not sure	22	23	21	11	10	15	13	12	10
2	Invalid	17	23	18	19	12	15	17	33	27
1	Very invalid	9	5	4	0	1	2	9	5	5
No of respondents		25	57	57	25	102	91	26	57	59
Mean		3.3	3.2	3.4	3.6	3.8	3.6	3.3	3.1	3.3
Standard deviation		1.1	1.0	1.0	0.9	0.9	1.0	1.1	1.1	1.2

[13] 96TQT4d, 96TQN2c, 97TQ2b, 98TQ2b
[14] 96TQT5b, 96TQN2b, 97TQ2a, 98TQ2a

Table 4.8 illustrates that each year at least half of the teachers who assessed against the PUS (52%-50%-58%) found the credits either "Valid" or "Very valid" for indicating achievement in Year 12 physics. Approximately a fifth of teachers (22%-23%-21%) were "Not sure" and about a quarter (26%-28%-22%) found the credits either "Invalid" or "Very invalid". There were no statistically significant longitudinal changes in this pattern of responses $(F\ (2, 136)_{1996/1997/1998} = 0.462, p > 0.05)$.

Overall the pattern of responses for SFC was similar. Each year at least half (61%-50%-58%) of the teachers who practised dual assessment and a higher percentage (70%-77%-68%) of teachers who assessed for SFC only felt that SFC grades were either "Valid" or "Very valid " for reporting student achievement in Year 12 physics. There were no statistically significant longitudinal effects $(F\ (2, 354)_{1996/1997/1998} = 0.36, p > 0.05)$. Teachers who assessed for SFC only $(M = 3.7, SD = 0.9)$ felt that SFC grades were significantly $(F\ (1, 354)_{PUS/SFC} = 11.61, p < 0.01)$ more valid than teachers who practised dual assessment $(M = 3.3, SD = 1.0)$. This was not unexpected because teachers who doubted the validity of SFC grades were probably more likely to trial assessment against the PUS.

There was no statistically significant difference between teachers' perceived validity of PUS credits and SFC grades $(F\ (1, 497)_{PUS/SFC} = 3.63, p > 0.05)$. Over the three years of the study, teachers who had assessed against the Level 2 PUS $(M = 3.3, SD = 1.0)$ rated the Level 2 PUS credits equally valid as teachers who assessed for SFC rated SFC grades $(M = 3.5, SD = 1.0)$.

The following comment illustrates the main reason given by teachers each year for considering PUS credits to be valid for describing students' achievement in Year 12 physics:

> The Unit Standard credits are a valid indicator of student achievement. The elements represent clear learning outcomes and the assessment schedules provide clear guidelines as to whether the students have achieved the element or not. The teacher's record shows exactly what

> areas of learning the students are competent in. This type of assessment avoids the problems associated with scaling and the linking of SFC grades with SC results.

Each year, teachers who felt that PUS credits were invalid for indicating students' achievement, referred to the system's inability to recognise partial or different levels of achievement and excellence. The following comment made in 1997 represents this view:

> There is a world of difference between a student who has narrowly failed to complete an element and one who has not attempted it. Yet both get the same recognition. Similarly there is a difference between completing an element first time and completing it after several reassessments. This removes all motivation to excel and only favours mediocrity.

Other teachers commented that the breaking up of the Year 12 course into units led to atomisation of learning outcomes that are better assessed in an integrated way:

> ... the compartmentalised discrete credits did not reflect the holistic nature of the subject. The way the Physics Unit Standards are set up tends to compartmentalise each topic and students don't get to see the link between different aspects of the physics course. Students who have achieved elements piecemeal may not have a very good overall understanding of the subject.

In contrast a teacher who thought that SFC grades were valid for indicating student achievement remarked:

> Sixth Form Certificate has been proven over nearly 20 years. It allows assessment of a good balance of practical and theoretical work and gives credit for partial success. The grade is a good indicator of general achievement and it is challenging for able students. All students who participate get a grade.

Another teacher commented:

> Sixth Form Certificate grades are an efficient mechanism for ranking students and establishing a hierarchy of grades and can reflect varying degrees of achievement and recognise and reward excellence.

Teachers who assessed for SFC, expressed the view that the SFC grades need to be more closely based on work done in Year 12 and not on a distribution of marks established by a means analysis of the previous year's SC results. Teachers also expressed dissatisfaction with the hierarchy of Year 12 subjects established by the grade allocation process because it leads to different subjects having different proportions of the top grades. One teacher complained:

> Students in non-academic subjects find it almost impossible to get a good grade even though they have a good understanding of the subject and in some cases have won national competitions. It is probably true to say that in an academic subject like physics, the poorer students get better SFC grades than they deserve. The historical hierarchy of subjects and the relative value of different kinds of knowledge and subjects are artificial and should be abolished.

Comments made by teachers from smaller schools warned that the statistical validity of the process of allocating grades may break down when applied to small classes. A representative comment was:

> The mechanism and statistical machinations for allocating SFC grades is invalid for small schools where you don't have a normal distribution. PUS credits report on what students have achieved whereas SFC grades are based on norm referenced marks in other subjects attained in the previous year. How valid is that?

Overall the above analysis indicates that each year at least half (50%-58%) of the teachers who had assessed against the Level 2 PUS found the credits valid for indicating student achievement in Year 12 physics. Furthermore, teachers perceived PUS credits to have similar validity to SFC grades for indicating student achievement in Year 12 physics. There were however positive and negative aspects associated with both systems of reporting.

Teachers liked reporting against the clear learning outcomes of the PUS but expressed concern about the system's inability to distinguish between different levels of achievement and excellence. On the other hand teachers liked the way SFC grades can be used to recognise different levels of achievement but expressed concern about the validity of the statistical moderation process, especially at smaller schools.

4.3.2 The transparency of the PUS reporting process

For a reporting process to be valid it must be transparent and understandable to students. If this is not the case, students may perceive it as unfair because it is hard to explain differences between results from different students in different classes or schools. Since larger schools generally have additional levels of inter-class moderation that may be difficult for students to understand, the effect of school size on students' understanding of the reporting process was also investigated.

Students' understanding of the reporting process for assessment against the Level 2 PUS was investigated two ways. The first was by surveying students who had been assessed against the Level 2 PUS. The second involved a comparison with students' understanding of the reporting process for SFC. [15] The results are displayed in Table 4.9. The SFC category includes students who were assessed for SFC only and students who were assessed for both SFC and the PUS.

[15] 96SQT6, 97SQ15, 98SQ11/96SQT10, 96SQN6, 97SQ7, 98SQ7

Table 4.9: Students' opinions on how well they understood the processes by which PUS credits and SFC grades were awarded

		Percentage of students					
		PUS			SFC		
		1996	1997	1998	1996	1997	1998
5	Very well	14	7	12	10	8	14
4	Well	42	31	41	40	49	42
3	Not sure	22	30	34	32	22	22
2	Poorly	17	21	9	13	16	17
1	Very poorly	5	10	4	5	4	5
No of respondents		375	220	196	830	481	501
Mean		3.4	3.0	3.5	3.4	3.4	3.5
Standard deviation		1.1	1.1	1.0	1.0	1.0	1.0

Table 4.9 shows that in 1996 and 1998, a small majority of students (56%-53%) understood the process by which credit was awarded for the PUS either "Well" or "Very well". However in 1997 only 38% of students responded this way. Each year, a minority (22%-31%-13%) of students understood the process "Poorly" or "Very poorly".

A two factor ANOVA revealed highly significant longitudinal differences *(F (2, 782)* $_{1996/1997/1998}$ *= 11.64, p < 0.0001).* The post-hoc Scheffé analysis revealed a decrease from 1996 to 1997 *(p < 0.001)* and an increase from 1997 to 1998 *(p < 0.001)*, in the mean level of student understanding of the process by which credit was awarded for the Level 2 PUS. The lower level of understanding in 1997 may be explained by the fact that NZQA support and provision of resources during 1997 was less than during the 1996 trial year. It might also have been affected by the unsettling effect of industrial action. Subsequent increases may be due to increased student familiarity with assessment against the PUS.

In comparison, each year approximately half to two-thirds (50%-57%- 66%) of students understood the process by which SFC grades are awarded either "Well" or "Very well" and only a small minority (17% -20%- 22%) of students understood the process "Poorly" or "Very poorly". There were no statistically significant longitudinal differences in students' understanding of the process of awarding SFC grades $(F\ (2,\ 1806)_{1996/1997/1998} = 0.44,\ p > 0.05)$.

An ANOVA to compare the levels of students' understanding of the processes by which PUS credits and SFC grades were awarded showed no statistically significant difference between students' understanding of the two processes $(F\ (1,\ 2601)_{PUS/SFC} = 2.04,\ p > 0.05)$. Students who were assessed against the Level 2 PUS $(M = 3.3,\ SD = 1.0)$, had a similar mean level of understanding of the process of awarding credits to students' understanding of the process by which grades were awarded for SFC $(M = 3.3,\ SD = 1.0)$.

The ANOVA also revealed that school size impacted significantly on students' understanding of the process by which credit was awarded ($F\ (2,\ 782)_{small/medium/large} = 7.98,\ p < 0.0005$). The post-hoc Scheffé analysis showed that students in large schools $(N = 317,\ M = 3.1,\ SD = 1.1)$ understood the process by which credit is awarded for the Level 2 PUS less well than students in medium $(N = 236,\ M = 3.5,\ SD = 1.1)$ and small schools ($N = 237$, $M = 3.5,\ SD = 1.0$). Both of these comparisons were significant at the $p < 0.001$ level. Possibly, smaller class sizes allow teachers more time to explain the reporting process to individual students. Furthermore, in larger schools there is generally more than one physics class and consequently a need for inter-class moderation. This process introduces another level of assessor judgements and may not be explained well to the students. In contrast there was no difference in students' understanding of the process by which SFC grades were awarded that was related to school size $(F\ (2,\ 1806)_{small/medium/large} = 0.28,\ p > 0.05)$. This difference in the effect of school size may be because the process of awarding credit involves resubmission and reassessment and consequently depends on a greater amount of teacher/student interaction than the process of awarding grades and is therefore more susceptible to class size factors.

Students who said they understood the process by which credits were awarded, either had the process explained well by their teachers or had received the NZQA leaflet *Explaining the difference* which provided information for Year 12 students on the differences between SFC and US. An obviously well informed student said:

> We got a good NZQA brochure to explain the system and our teacher explained the whole system thoroughly. I understand the assessing, reassessing and credit accumulation and the performance criteria are largely self explanatory. You can either get the Unit Standard and can do the work or you can't. So you get the credit or you don't.

A minority of students complained that they had not received any information from NZQA. Matheson (1997: 5) said that a number of principals withheld the NZQA leaflet because they felt that it:

> ... would give the impression to students and their parents that Sixth Form Certificate is not worth the paper it is written on.

There were three aspects that were poorly understood by students. The first was uncertainty about what qualification the credits gained by being assessed against PUS count towards. A typical student comment was:

> The Unit Standard credits have not been explained properly. What qualification do they count for? I don't know how they work out how many credits you get for each Unit Standard test and don't understand how many units are needed for a certificate or what the certificate is called.

Secondly students were unsure about the US requirements for entry into Year 13 or tertiary study. One student stated:

> I don't know how many credits you need to get into the Year 13 Physics course. I want to do engineering at university. Does the university accept the Unit Standards?

The third aspect related to lack of transparency in how teachers decide whether a student is "competent" or "not yet competent" for a particular element. A typical response was:

> It is difficult to see how the performance criteria relate to the way the tests are marked sometimes. You get a tick for what you did right but the teacher doesn't identify what you did wrong and why.

This is probably due to the fact that students are supplied with the performance criteria but do not have access to the judgement statements in the assessment schedule. These interpret the performance criteria for the specific context of the assessment activity.

Students' understanding of the process of awarding SFC grades was affected by similar factors to those listed for the PUS. A typical comment made by students who understood the process by which SFC grades were awarded was:

> The process was well explained. We were all told at the start of the year, things were explained in a newsletter, much the same as other subjects. It is quite straightforward.

Typical comments made by students who did not understand the process included:

- Seriously I don't think anyone in the class has any idea of how the Sixth Form Certificate grades are arrived at.
- I have not had it explained it is a bit of a mystery.
- What percentage does a test, exam or practical contribute towards your final grade?
- I am not clear as to how the grade pool works.

The practice of dual assessment caused some confusion for students. One student thought that:

> You are awarded a Sixth Form Certificate grade that depends on how many Unit Standard credits you have achieved.

Based on these findings, it may be concluded that the process of awarding credit was reasonably transparent to students and similar to students' level of understanding of the way SFC grades were awarded. A factor that needs to be taken into account however was that students from small and medium schools had a better understanding of the process than students in large schools.

4.3.3 Usefulness of Year 12 physics reporting for describing student achievement

In addition to satisfactory validity of credits and a transparent reporting process, it is essential that the results of the assessment against the PUS are a useful description of students' achievement in Year 12 physics. The following section discusses teachers' and students' opinions on this issue. A comparison is made with reporting for SFC.

4.3.3.1 Teachers' views

Each year of the study, teachers who had assessed against the Level 2 PUS were asked how useful they found the performance criteria to indicate student mastery of the learning outcome stated in the element[16]. Table 4.10 contains a summary of their responses.

[16] 96TQT2b, 97TQ14, 98TQ13

Table 4.10: Teachers' opinions on the usefulness of the performance criteria for describing student achievement (1996-1998)

		Percentage of teachers		
		1996	1997	1998
5	Very useful	13	10	9
4	Useful	74	62	61
3	Not sure	0	14	15
2	Of limited use	13	12	15
1	No use	0	2	0
No of respondents		26	49	54

($\chi^2_{1996/1997/1998} = 4.84$, $df = 6$, $p > 0.05$)[17]

Table 4.10 shows that each year a substantial majority (87%-72%-70%) of teachers found the performance criteria either "Useful" or "Very useful", and only about 15% found them "Of limited use" or "No use". In 1997 and 1998 a small minority (14%-15%) were "Not sure". There were no statistically significant longitudinal effects.

Teachers who found the performance criteria useful referred to the diagnostic nature of standards-based assessment. One teacher commented:

> The records of the performance criteria that have been achieved showed up exactly what skills students have mastered, and highlighted specific areas where individual students or most of the class has missed the point and this enabled me to go back and revise those aspects at the student and or class level.

Teachers who found the criteria to be "Of limited use" (13%-12%-15%) remarked that:

> The performance criteria in the Physics Unit Standards are very prescriptive and specific and lead to a high degree of reliability in

[17] categories 1 and 2 were combined.

> assessor judgements and ease of marking. The consequence of this is that students often fail to attain the element for presenting an answer in an unexpected format. It also makes it difficult to be creative when designing assessment activities.

A common criticism was that the performance criteria were too pedantic. One teacher commented:

> The performance criteria are too specific and picky and small silly mistakes often prevent students from achieving credit even though you feel that they have the right idea and a good understanding of the principles.

Examples of this identified by teachers were:

> ... performance criteria that focus on aspects such as stating the answer in the form of a complete sentence or to a specific number of significant figures.

Another criticism was that some performance criteria were considered ambiguous and that others could not be interpreted without referring to the specific judgement statements contained in the marking schedules of the sample assessment activities contained in the *Assessment Guide: Physics*.

It was felt that some PUS required too many performance criteria for gaining credit, and that there was too much repetition of the same performance criteria in several PUS, particularly the showing of working. The following remark highlights this repetition:

> If a student gains credit for finding an answer and showing their working in Physics Unit Standard 6380[18], they may also have to meet similar performance criteria in another Physics Unit Standard. I find this too fussy and repetitive.

[18] *Apply formulae, graphical and vectorial methods to find unknowns for a physical system.*

4.3.3.2 Students' views on the usefulness of PUS assessment results for describing achievement

Each year of the study students were asked how useful they found the results and feedback of their assessment against the PUS/SFC for describing how well they performed on a particular task.[19] The results are summarised in Table 4.11.

The PUS results included information on which elements and individual performance criteria a student had achieved but teachers were advised at the cluster meetings not to give written feedback that identified specific aspects for future improvement because this could invalidate students' opportunities for resubmission. The results of assessments for SFC generally included a percentage or letter grade and written comments.

[19] 96SQT7, 97SQ16, 98SQ12 / 96SQT11, 96SQN7, 97SQ8, 98SQ8

Table 4.11: Students' opinions on the usefulness of results of assessment against the PUS and SFC for describing how well they performed on a particular task (1996-1998)

		Percentage of students					
		PUS			SFC		
		1996	1997	1998	1996	1997	1998
5	Very useful	5	3	8	10	15	15
4	Useful	24	38	36	46	47	48
3	Not sure	29	24	28	25	18	20
2	Of limited use	27	23	20	15	15	15
1	No use	15	12	8	4	5	2
No of respondents		372	218	196	849	507	494
Mean response		2.8	3.0	3.2	3.4	3.5	3.6
Standard deviation		1.1	1.1	1.1	1.0	1.1	1.0

In contrast to teachers' endorsement of the usefulness of PUS assessment results, Table 4.11 indicates that students' opinions were divided. Each year a sizeable minority (29%-41%-44%) of students felt that the results of assessment against the level 2 PUS were either "Very useful" or "Useful" for describing how well they performed on a particular task. A similar sized group (42%-35%-28%) felt they were "Of limited use" or "No use" and about a quarter (29%-24%-28%) were "Not sure". There were no statistically significant longitudinal changes in students' opinions $(F(2, 777)_{1996/1997/1998} = 0.11, p > 0.05)$.

In contrast to this students felt that SFC assessment results and feedback served this purpose better. Each year a majority (56%-62%-63%) of students felt that the marks and comments on assessments for SFC were either "Useful" or "Very useful" for describing how well they performed on a task and only about a fifth felt that they were "Of limited use" or "No use".

A single factor ANOVA showed that over the three years of the study there was a significant difference between students' ratings of the usefulness of assessment results for SFC and the PUS for describing how well students performed on a particular task ($F\ (1, 2634_{PUS/SFC} = 116.71, p < 0.001)$. Students who were assessed for SFC rated assessment results and feedback ($M = 3.5$, $SD = 1.0$) more useful for describing achievement than students who were assessed against the Level 2 PUS rated PUS assessment results ($M = 3.0$, $SD = 1.0$).

The reason students gave most frequently for not finding the PUS results useful for describing how well they performed related to the nature of competency-based assessment. A typical response was:

> The credits tell you whether you have passed or not but they don't tell you how well you have done. I don't like the fact that I can never do better than a pass.

In contrast students felt that the marks and comments they received on their assessments for SFC were:

> a good indicator of the overall level of achievement in Year 12 physics. The grades are not specific but tell you about your general knowledge of physics. You can monitor your progress and keep track of where you are going. Grades may not pinpoint what we don't know as the Unit Standard tests do but at least we get recognition of what we have achieved. It is not a pass/fail system.

A two factor ANOVA revealed highly significant differences in students' perceptions of the usefulness of PUS results that related to school size ($F\ (2, 777)_{small/medium/large} = 10.71, p < 0.0001$). The post-hoc Scheffé analysis revealed that students in large schools ($N = 317$, $M = 2.6$, $SD = 1.1$), found the results of assessment against the Level 2 PUS significantly less useful ($p < 0.001$) for describing how well they performed on a task than students in medium ($N = 236$, $M = 3.1$, $SD = 1.1$) and small schools ($N = 233$, $M = 3.1$, $SD = 1.0$).

The SFC data showed a parallel trend ($F\ (2,\ 1844)_{small/medium/large} = 47.82,\ p < 0.0001$). The post-hoc Scheffé analysis revealed that students in small schools ($N = 374,\ M = 4.0,\ SD = 1.1$) found the results of assessment for SFC significantly more useful ($p < 0.001$) for describing how well they performed on a task, than students in medium ($N = 691,\ M = 3.4,\ SD = 1.0$) and large schools ($N = 785,\ M = 3.5,\ SD = 1.0$).

A possible explanation of the effects of school size on students' perception of the usefulness of both SFC and PUS assessment results is that small schools generally have smaller classes that enable teachers to spend more time with individual students. In the case of SFC this may manifest itself in more comprehensive written and verbal feedback on assessment activities. In the case of the PUS, teachers are able to spend more time on processing resubmissions and reassessments. Furthermore students in smaller schools with fewer and smaller physics classes have a clearer overview of the range of students' ability in physics and where they personally fit into that range. This understanding may make the assessment results appear more meaningful.

The investigation into the usefulness of PUS results for describing students' achievement resulted in the following main research findings.

Over 70% of teachers found the performance criteria useful for describing students' achievement. A single factor ANOVA revealed a highly significant difference between teachers' and students ' opinions on the usefulness of PUS assessment results for describing achievement ($F\ (1,\ 913_{Teachers/Students} = 46.13,\ p < 0.001$). Students found the performance criteria ($N = 129,\ M = 3.7,\ SD = 0.850$) significantly less useful for describing achievement than teachers did ($N = 786,\ M = 3.0,\ SD = 1.1$) and would prefer additional targeted written feedback. A possible explanation for this is that teachers value the diagnostic value of the reporting process for the PUS whereas students prefer the more competitive norm-referenced nature of the reporting process for SFC. This is closely linked to the debate about the recognition of excellence in the PUS reporting process.

4.3.4 The recognition of excellence

Assessment against the PUS is competency-based. If students demonstrate competence for particular PUS, they receive the credit and move on to another PUS, either at the same level or at a higher level. If students do not meet the requirements of the PUS, they are judged to be "not yet competent" and need to be reassessed at a future time. In the 1996 questionnaire teachers remarked that the competent/not yet competent distinction does not acknowledge different levels of achievement and fails to recognise excellence. These comments were further explored by the 1997 questionnaire. Physics teachers were asked whether they thought assessment against the PUS enabled recognition of excellence.[20] Sixty-eight percent of teachers felt that assessment against the PUS did not enable recognition of excellence and 32% felt that it did.

Teachers who felt that assessment against the PUS did not enable recognition of excellence argued that since the assessment is competency-based it does not recognise achievement that surpasses the standard. This reasoning is reflected in the comment:

> Some students pass the Unit Standard on the first attempt, others after two or three reassessments, yet they both receive the same number of credits. Their respective Records of Learning show no difference. Therefore assessment against the Physics Unit Standards doesn't recognise the extra problem solving ability of the more able students.

They also felt that a rank order of test results is needed for students to compete and strive for excellence.

[20] 97TQ21

Teachers who answered affirmatively said that the NQF does recognise excellence because the ROL reflects excellence by the increase in the number and levels of the credits registered. As one teacher commented:

> More able students accumulate more credits and at higher levels. The excellent can attain significant numbers of units or a higher level National Certificate.

Other ways of recognising excellence recommended by teachers included comments and grades on school reports and internal school awards and prizes.

In 1997 and 1998, all teachers in the study were asked whether they felt that criteria for excellence should be built into each PUS.[21] Table 4.12 summarises their responses.

Table 4.12: Teachers' opinions on whether criteria for excellence should be built into the PUS (1997-1998)

		Percentage of teachers	
		1997	1998
5	Strongly agree	44	34
4	Agree	33	43
3	Not sure	9	7
2	Disagree	8	10
1	Strongly disagree	6	6
No of respondents		154	142

($\chi^2_{1997/1998}$ = 4.69, df = 4, p > 0.05)

Table 4.12 indicates that in 1997 and 1998, approximately three-quarters (77%) of teachers either "Agreed" or "Strongly agreed" that criteria for excellence should be built into the PUS and only a small minority (14%-16%) disagreed with this view. The small percentage of teachers who were "Not sure" (9%-7%) indicates that teachers had generally made up their minds on

[21] 97TQ4d, 98TQ4f

this issue. There were no statistically significant differences between 1997 and 1998.

A typical comment made by teachers who supported the incorporation of criteria for excellence was:

> The lack of recognition of excellence is a major failing of this system. We need to generate excellence. We need to know the difference between someone mastering the skill and someone with outstanding skills. At present the Physics Unit Standards cater only for the mediocre. Students with exceptional ability are not recognised.

A representative comment made by teachers who did not agree with the inclusion of criteria for excellence was:

> Additional criteria for excellence are not necessary. If you are going to grade you might as well go back to the old system of exams. There is scope in the present system for students to do more Unit Standards or move to higher levels in the framework. The complexity of the system is a major drawback. Making it more complex would not be a good thing.

In 1997 teachers were asked if they had any suggestions about how criteria for recognising excellence could be built into assessment against the PUS.[22] Teachers offered the following suggestions:

- Award bonus credit for achieving the Unit Standard at the first attempt.
- Introduce two levels of achievement for each element. These levels could be competency and merit.
- Award credit for achieving part of a Unit Standard, perhaps at the element level.
- The National Certificate could be awarded with honours for extra credits at a certain level.

[22] 97TQ22

- A Unit Standard could be achieved with guidance for competence and without guidance for excellence.
- Include range statement of two or three items, like displacement, velocity and acceleration. Competence, merit or excellence could be awarded depending on the number of items the student got correct.
- Schools could recommend students who excelled for recognition by NZQA.

4.3.5 The diagnostic value of PUS assessment results

In addition to being useful to describe student achievement, assessment results must serve a diagnostic function and provide guidance for students on how future achievement may be improved. The most appropriate group to judge whether this criterion had been achieved, were the students who had been assessed against the Level 2 PUS. In addition a comparison was made with students' perception of the diagnostic value of SFC assessment results. Each year of the study students were asked how useful they found the results they received on their assessments against the PUS and SFC for describing how they could improve their future performance.[23] Table 4.13 shows the results for students who were assessed for SFC[24] and students who were assessed against the Level 2 PUS.

[23] 96SQT8, 97SQ17, 98SQ13 / 96SQT12, 96SQN8, 97SQ9, 98SQ9
[24] The SFC samples include students who were assessed for SFC only and students who were assessed for both SFC and the PUS.

Table 4.13: Students' opinions on the usefulness of results of assessment against the PUS/SFC for describing how performance can be improved (1996-1998)

		Percentage of students					
		PUS			SFC		
		1996	1997	1998	1996	1997	1998
5	Very useful	7	5	7	10	12	13
4	Useful	29	31	29	35	46	41
3	Not sure	33	32	30	27	23	26
2	Of limited use	16	19	21	17	14	17
1	No use	15	13	13	11	5	3
No of respondents		352	206	189	848	506	491
Mean response		2.9	2.9	3.1	3.2	3.5	3.4
Standard deviation		1.2	1.1	1.1	1.1	1.0	1.0

Table 4.13 shows that students' opinions on the usefulness of results of assessment against the Level 2 PUS for describing how performance could be improved were divided. Each year 36% of students felt that the results were "Useful" or "Very useful", approximately a third (33%-32%-30%) were "Not sure" and a further third (31%-32%-34%) felt that they were "Of limited use" or "No use" for this purpose. There were no statistically significant longitudinal differences $(F\ (2,\ 738)\ _{1996/1997/1998} = 0.29,\ p > 0.05)$.

In comparison, a more substantial percentage of students (45%-58%-54%) each year found the results of assessment for SFC "Useful " or "Very useful" for describing how their performance could be improved. About a quarter of students (27%-23%-26%) each year were "Not sure" and only a minority (28%-19%-20%) felt that the results were "Of limited use " or "No use". There were statistically significant longitudinal differences between the mean views $(F\ (2,\ 1839)\ _{1996/1997/1998} = 12.35,\ p < 0.001)$. The post-hoc Scheffé analysis showed that the statistically significant increases occurred between 1996 and1997 *($p < 0.001$)* and between 1996 and 1998 *($p < 0.001$)*.

A single factor ANOVA *($F(1, 2590)_{PUS/SFC} = 54.35$, $p < 0.0001$)* showed that students who had been assessed for SFC found their assessment results more useful *($M = 3.3$, $SD = 1.1$)* for describing how performance could be improved than students who had been assessed against the Level 2 PUS *($M = 3.0$, $SD = 1.1$)*.

Students who felt that the feedback they received on the PUS assessment activities was useful, commonly made reference to resubmission and reassessment. The following comments are representative:

- It shows in what performance criteria you have met and which performance criteria are incomplete. This is useful in directing my study and in preparation for reassessment.
- You can work out where you went wrong. It highlights what you don't know and helps you plan your revision for reassessment and directs your learning efforts.
- The Unit Standard results tell you what you can do and what you can't do. It helps you to direct your efforts and tells me what points I need to improve on and specifies what you need to do to pass.
- Tells me what I have to do at a later date if I want to resubmit.

A possible reason for student dissatisfaction with the diagnostic value of the assessment results may be that teachers were instructed that the feedback should not be specific because the students are offered the opportunity to resubmit their work. If the feedback was too specific it could provide too much guidance and make the resubmission process invalid.

Students who felt that the feedback was not useful complained about the lack of written teacher feedback. One student remarked:

> A tick or cross in a checkbox was not sufficient or specific enough. A comment on how you might go about getting it right next time so you have to work it out yourself would be much better.

In comparison, students thought that the feedback on assessments for SFC was useful because it had diagnostic value. The end-of-unit marks tell the students which unit they understand well and which topics they need to improve on. One student remarked:

> I find it useful because it is an indication of how I am doing and it highlights the parts that I don't understand so that I can focus more on those parts.

Another student commented:

> It's really helpful to know how well you did, so you can work on things that need it. It sort of helps me to understand my strengths and weaknesses in this subject.

Another group of SFC student responses highlighted that in addition to the marks, the written comments made by the teacher provided useful additional feedback . A typical response in this category was:

> The feedback from the teacher is good. He goes over the work thoroughly and provides model answers to show you where you went wrong.

A two factor ANOVA revealed highly significant differences in students' perception of the usefulness of the diagnostic value of PUS assessment results that related to school size $(F(2, 738)_{small/medium/large} = 14.36, p < 0.0001)$. The post-hoc Scheffé analysis revealed that there was a significant difference between large and medium $(p < 0.01)$ schools. There was also a highly significant difference between large and small schools $(p < 0.001)$. Students in large schools ($N = 305, M = 2.8, SD = 1.2$) found the written feedback they received less useful than students in medium ($N = 219, M = 3.1, SD = 1.1$), and small schools ($N = 223, M = 3.3, SD = 1.0$).

Analysis of the SFC data revealed a similar trend. A two factor ANOVA revealed significant differences related to school size

($F(2, 1836)_{small/medium/large} = 14.13, p < 0.05$). The post-hoc Scheffé analysis revealed that there were significant differences between large and small ($p < 0.05$) and medium and small schools ($p < 0.05$). Students in small schools ($N = 374, M = 3.4, SD = 1.1$) found the results of assessment for SFC more useful for describing how they could improve their performance than students in medium ($N = 689, M = 3.3, SD = 1.1$) and large schools ($N = 782, M = 3.3, SD = 1.1$). A possible explanation for this is that smaller class sizes allow teachers to give more comprehensive feedback.

The overall conclusion is that students were not sure about the value of results of assessment against the PUS for describing how their future performance could be improved but that they found the results less useful than the results of assessments for SFC. Students in smaller schools found both SFC and PUS assessment results more useful than students from larger schools.

4.3.6 Conclusion about the validity of reporting of PUS assessment results

The investigation into the validity of the reporting process of assessment against the Level 2 PUS produced the following research findings:

- The majority of teachers found the Level 2 PUS credits valid for reporting achievement in Year 12 physics.
- The process of awarding credit was understood by the majority of students and was similar to students' level of understanding of the process of awarding SFC grades.
- Each year at least 70% of teachers found the results of assessment against the Level 2 PUS useful for describing students' achievement. Students found the results substantially less useful than teachers did and also less useful than SFC assessment results.
- A majority of teachers felt that the reporting process for assessment against the Level 2 PUS did not enable recognition of excellence and that criteria for excellence should be built into the PUS.

- Students were divided in their opinions about the diagnostic value of the PUS and found SFC grades more useful for this purpose because of the targetted written feedback.
- Students from small schools had a better understanding of the reporting process and found the assessment results more useful for describing performance and how it could be improved than students from larger schools.

4.4 Consequential validity of assessment against the PUS

In addition to high curriculum fidelity, concurrent validity and validity of reporting, it is necessary for assessment against the Level 2 PUS to have satisfactory consequential validity. Consequential validity was addressed by the research question:

What is the impact of assessment against the Level 2 PUS on students and teachers?

The investigation focused on the following areas of impact:

- students' satisfaction with the way they were assessed
- student learning
- student motivation
- student enjoyment of the Year 12 physics course
- teacher enthusiasm
- classroom teaching.

These are discussed in detail in the following sections.

4.4.1 Students' level of satisfaction with the way they were assessed

One aspect of consequential validity investigated was how satisfied students were with the way they were assessed against the PUS and how this compared with students' level of satisfaction with assessment for SFC.[25] Table 4.14 summarises students' responses.

Table 4.14: Students' level of satisfaction with the way they were assessed for SFC and against the PUS (1996-1998)

		Percentage of students					
		PUS			SFC		
		1996	1997	1998	1996	1997	1998
5	Very satisfied	4	6	8	6	8	6
4	Satisfied	21	30	39	53	48	54
3	Not sure	33	24	26	23	23	28
2	Unsatisfied	26	27	18	13	16	7
1	Very unsatisfied	16	13	9	5	5	5
No of respondents		376	213	193	849	507	496
Mean response		2.7	2.9	3.2	3.4	3.4	3.5
Standard deviation		1.1	1.1	1.1	1.0	1.0	0.9

Table 4.14 shows that over the three year period, a growing minority of students (25%-36%-47%) were either "Satisfied", or "Very satisfied" with the way they were assessed against the Level 2 PUS. There was a corresponding decreasing trend (42%-40%-27%) in the percentage of students that were "Unsatisfied" or "Very unsatisfied". An ANOVA revealed significant longitudinal differences in the mean level of students' satisfaction with assessment against the PUS. *($F(2, 773)_{1996/1997/1998} = 3.80$, $p < 0.05$)*. The post-hoc Scheffé analysis showed that student satisfaction increased significantly from 1996 to 1998 *($p < 0.001$)*.

In contrast a larger and steady percentage (59%-56%-60%) of students were "Satisfied" or "Very satisfied" with the way they were assessed for SFC over the period of the study. About a quarter (23%-23%-28%) of students were

[25] 96SQT8, 97SQ12, 98SQ10

"Unsure" and a small minority (18%-21%-12%) was "Unsatisfied" or "Very unsatisfied". There were no statistically significant longitudinal differences $(F(2, 1840)_{1996/1997/1998} = 1.88, p > 0.05)$.

A single factor ANOVA revealed that over the three years of the study, there was a highly significant statistical difference between students' levels of satisfaction with the way they were assessed for SFC and the PUS $(F(2, 2603)_{PUS/SFC} = 158.04, p < 0.001)$. Students who were assessed for SFC were more satisfied with the way they were assessed $(M = 3.4, SD = 1.0)$ than students who were assessed against the Level 2 PUS $(M = 2.9, SD = 1.1)$.

The ANOVA also revealed highly significant differences related to school size $(F(2, 773)_{small/medium/large} = 20.07, p < 0.0001)$. The post-hoc Scheffé analysis revealed that there were significant differences between large and medium ($p < 0.001$) and large and small schools ($p < 0.001$). Students in large schools $(N = 317, M = 2.6, SD = 1.1)$ were less satisfied with the way they were assessed against the Level 2 PUS than students in medium ($N = 234, M = 3.1, SD = 1.1$) and small schools ($N = 231, M = 3.1, SD = 1.1$). There were no statistically significant differences in student satisfaction with the way they were assessed for SFC that related to school size $(F(2, 1840)_{small/medium/large} = 1.22, p > 0.05)$.

Each year the quantitative investigation into students' level of satisfaction with the PUS was supplemented by asking students a more open-ended question about what aspects of assessment against the PUS they liked or disliked.[26] The responses were similar each year and were collated under headings that are summarised in Table 4.15. The headings are not presented in order of frequency of responses. The aim was not to provide an indication of strength of feeling on individual issues, but to illustrate the full range of issues raised by students in order to add context to the quantitative data presented in Table 4.14. Students' comments frequently drew comparisons with assessment for

[26] 97SQ13-14, 98SQ20-21

SFC. These are incorporated into the discussion to provide a basis for comparison and contrast.

Table 4.15: Year 12 physics students' views on assessment against the PUS

Aspects of assessment against the PUS that Year 12 physics students:	
liked	disliked
• Explicit learning outcomes • Lower stakes assessment close to learning • Credit at the unit level • Reassessment • Absence of scaling • Flexible distribution of credit • Comparability between schools and classes • Diagnostic and formative assessment • Facilitation of thorough learning • Independence of subjects • Seamlessness of the NQF	• The all or nothing reporting of outcomes • The pedantic nature of the performance criteria • Failure to recognise excellence • No differentiation between levels of achievement • Lack of recognition of partial understanding • Lack of competition • Disproportionate allocation of credits • Lack of acceptance by employers and tertiary institutions • Over-assessment • Inconsistencies between schools • Difficulty of achieving credit

The following sections discuss the students' comments in greater detail.

a) Aspects of assessment against the PUS that students liked

Each year students identified the following aspects of assessment against the Level 2 PUS that they liked:

Explicit learning outcomes

Students commented that the elements in the PUS clearly spelt out what could be assessed and that this made it easier to prepare for tests. One student remarked:

> The elements and criteria make it clear what is expected. You know exactly what to concentrate on when you are revising for a test. With SFC you never know exactly what is going to be examined and sometimes the tests are not relevant to what we have done in class.

Lower stakes assessment close to learning

Students liked the concept of being assessed throughout the year with a larger number of smaller lower stakes assessment activities. One student explained:

> There is not a big exam at the end of the year about everything you have ever done. The tests immediately follow each set topic. It makes the passing off the topic easier because the material is fresh in your mind. Smaller and more frequent tests help to take the pressure of because there is less importance placed on each test.

The assessment programmes for SFC generally consist of a smaller number of longer and higher stakes assessments. One student who was assessed for SFC commented remarked:

> We have a couple of big tests during the year. By the time we sit each test we have moved on to new topics and have sometimes forgotten earlier work. There is also too much emphasis on each test.

Credit at the unit level

Students liked receiving credit at the completion of each PUS and the way they could accumulate credit towards the National Certificate. They also liked the way each unit provided them with a "fresh start" and that failure in one unit did not jeopardise their chances for credit in another PUS. In contrast students who were assessed for SFC did not like the way the individual test marks were accumulated into a final SFC Grade. The final SFC grade might be a failing grade even though the student had achieved good marks in a number of individual unit tests. Students expressed the opinion that they should receive credit for what they can do well, rather than have this hidden by a final failing grade. As one typical SFC candidate put it:

> With only 4 common tests determining our marks all we have to do is stuff up on one and our grades go down considerably. The good marks in one topic can be cancelled out by a bad mark in others. Unit Standards are better because each unit is judged on its own merit and gives you a fresh start.

Reassessment

Students were generally given one or two reassessment opportunities and saw this as an advantage over assessment for SFC. One student referred to reassessment as:

> … being given a second chance and an opportunity to learn from our mistakes because if I don't get the standard first time I am given the chance of doing it again and correcting my mistakes and not lose any marks.

Students who underwent dual assessment and had been introduced to reassessment as part of their assessment against the PUS felt that this aspect was missing from assessment for SFC. As one student put it:

> With Sixth Form Certificate you don't get a second chance. If you fail you can't be re-assessed. We move on to the next topic and you don't get an opportunity to show that you have learnt from your mistakes. What is the point of sitting a test if you don't get a chance to show that you have improved?

There was a commonly held view by SFC candidates that competition for predetermined grades was the focus of the process by which SFC grades are awarded. One student said:

> There is too much pressure to compete with everyone else in your class as well the other classes. Your grade depends on your place in class, not on how much physics you know.

Absence of scaling

A large number of students liked the absence of scaling and the transparent nature of assessment against the PUS. A typical response was:

> You get credit if you meet all the requirements. The results are not scaled and there is no limit on the number of students who are allowed to receive credit. The Physics Unit Standards' assessments tell you what aspects of physics you know and what you don't know.

In contrast, the most commonly given reason for being dissatisfied with SFC was that the marks are scaled. As one student said:

> I don't like scaling. The marks I earn I should keep. Scaling is stupid, frustrating and confusing. It is unfair to have your marks scaled to someone else's School Certificate marks from the previous year.

SFC students commented that scaling sometimes distorted marks and that their marks did not necessarily reflect their ability in the subject. A typical comment was:

> If I get a good mark, it sometimes ends up lower than my bad marks by the time they are scaled up. There doesn't seem to be much credit for hard work. You should be given the average of your actual test marks for the year and not a scaled grade. If you deserve 70% you should get it, not a lower grade because you are under other people.

Flexible distribution of credit

Students liked the fact that there is no predetermined distribution of credits which 'condemns' a proportion of the class to fail, before they have even sat the assessment. One student commented that:

> The credits are not scaled and are not dependent on the distribution of marks based on the previous year's work. The SFC system of limiting the number of "ones" per school sucks. There is no limit on the number of students who can gain credit for particular PUS and there is no built in compulsory failure rate.

Students felt that the number of high grades should not be limited and that a grade should reflect achievement in physics not the students' rank order in class. The following comment was typical:

> I don't like the way the grades are limited. If you get over say 85% you should get a one. The grades should relate to how well you know physics not the SC results of other people. I think it is a very, very unfair system because there are only a certain number of students who will be able to attain good grades.

Comparability between schools and classes

Students commented that because the PUS contained clear and specific learning outcomes, all schools that offered the same PUS had to cover the same material. They felt that this approach introduced comparability between schools. This was in contrast to SFC where a number of students commented that they knew people in other schools or classes who did quite different things in their physics course.

Diagnostic and formative assessment

Students liked the fact that the PUS assessment results were diagnostic and provided the students with exact details of what they did and did not know. This coupled with the reassessment opportunities motivated and directed further learning. One student explained:

> If I don't understand something, the Unit Standard assessments help me to pinpoint exactly what I need to revise and I can pick up the credit on the reassessment. They help you to get better reassessments and encourage you to have another go.

They were concerned however at not receiving detailed written feedback on their work that showed them where they went wrong. In contrast students indicated that SFC assessments were summative and did not allow reassessment. A typical responses was:

> The test are there mainly to collect marks that count towards SFC not to guide you in your learning

Some students who were undergoing dual assessment made the comparison between the two different forms of assessments. A typical comment that illustrates this was:

> The marks don't tell you the areas in which you need to improve. They are just numbers, not descriptions like the performance criteria of the Unit Standards.

Facilitation of thorough learning

A small number of students indicated that the PUS encouraged thorough learning because the assessments highlighted what they did not know and gave them reassessment opportunities to address what they did not understand. One student remarked:

> The Unit Standards are very useful because to get the credit you have to know the work. I felt encouraged to persevere and really get to know a topic. You feel that if you get the credit you really understand the work unlike SFC where I sometimes get a good mark even though I do not really understand the work.

Independence of subjects

Students liked the way each subject was treated independently and that units of equal credit contributed equally to the National Certificate. This was in contrast to SFC where students felt that their final grade was influenced by students' performance in other subjects and even previous years. One student perceived that:

> Why should the marks a teacher is allowed to give in one subject be determined by anything else but the achievement of the students in the class and how well they know the subject.

Seamlessness of the NQF

The PUS credits that students achieve at school are building blocks for qualifications which students can complete after they leave school. A number of physics students commented that they had completed some electro-technology US as part of their physics course. This was seen to bring more relevance to school programmes, particularly by students who wanted to continue their study at a polytech or a private training establishment.

b) Aspects of assessment against the PUS that students disliked

Each year, students identified the following aspects of assessment against the Level 2 PUS that they disliked:

The all or nothing reporting of outcomes

Students did not like the "competent/not yet competent" nature of assessment against the PUS and the fact that the system does not recognise partial completion of performance criteria. A large number of students referred to the specific nature of the performance criteria and the fact that if they failed to present evidence for one performance criterion, they had to be reassessed for the whole element. A typical comment was:

> I don't like the fact that you need to get 100% correct to get credit for a US. If you make a small mistake in a specific part of a test you don't get credit, even though you may have done quite well overall.

This all or nothing approach had a negative impact on student motivation and enthusiasm for assessment against the PUS, and is reflected in the comment:

> It is very discouraging because you lose whole units just for a few mistakes. You should get marks where marks are deserved.

The pedantic nature of the performance criteria

The detail required by some performance criteria was seen by many students as "too picky". One student quoted a specific example:

> I missed out on a US even though I got the whole problem correct and understood it but because I wrote "sec" instead of "s" I failed. I don't like the way you can have an answer so near but you can't get a US, every detail has to be correct to attain the standard.

In addition, the judgement statements in the assessment schedule which link the performance criteria of an element to the evidence required in a specific assessment activity were seen as too specific by some students. One student said:

> It is pretty shaky to decide what's a pass and what's a fail in some cases, all it takes to pass is to say the sentence in a different way or include a key word etc. when you do actually understand it. Our teacher has marked us very strictly and has made our assessment very difficult to attain.

This was contrasted with norm-referenced assessment where students said you "got some recognition for getting things partly right".

Failure to recognise excellence

More able students expressed the view that assessment against the PUS did not recognise nor reward excellence of achievement. One student said:

> I think it is too easy to get them and then you cannot distinguish the top students from those with less ability.

Some expressed the view that there should be a mechanism whereby students could receive extra recognition for performance that exceeds the requirements of the PUS. One student complained:

> ... there is no way of being awarded for over average performance in individual credits. Why can't this be recorded on a Record of Learning?

At the other end of the achievement spectrum was the concern that the ROL of a student who attained a PUS on the first attempt, registered the same number of credits, as a student who met the requirements after a number of re-submissions on some elements and reassessment on others. A response that illustrated this was:

> There is no distinction between those students who passed the standard the first time around and those who passed it on their first, second or even third reassessment.

In comparison, students who were assessed for SFC liked the capability of SFC grades to reflect and report excellence. One student said:

> I like the competitive aspect of Sixth Form Certificate. If my marks are better than other students in my class, my Sixth Form Certificate will show this. This is important to me because when I start competing for jobs after I leave school, I want my qualifications to show that I am better than someone who has only just passed.

No differentiation between levels of achievement

Closely related to the criticism of the all or nothing approach was student feeling that someone who narrowly missed the requirements of a US should receive recognition that their performance was better than someone who missed out on a number of elements. One respondent said:

> No credit is given if you almost get the standard. You are no different to someone who has failed really badly. Neither do you get any recognition for surpassing the requirements.

In contrast to this students liked the way the results of SFC assessments enabled recognition of a wider range of different ability than the two-point scale used for the PUS and can reflect excellence. One student remarked:

> I believe that my marks reflect my ability in physics and feel that the tests are fair. With the Unit Standards you can get very close to a pass and still not get recognition for your achievement. I feel this is unfair.

Lack of recognition of partial understanding

Students did not like the fact that they might have met nearly all the performance criteria for a PUS but did not receive any recognition for it. They preferred SFC where percentage grades could give an indication of partial achievement instead of the not yet competent nature of US reporting. One student explained:

> You don't have to get a 100%; your grade shows how much of the topic you know. Even 45% is better than a 'fail'.

Students also liked the way the percentage results for individual questions are added up to give a global percentage result for the test. This way high mark questions compensate for the low mark questions and when added up can still lead to a passing grade (50% was commonly mentioned as a pass). One student said:

> I find it very demanding and consistent yet it is very fair and gives more chance for people to get over 50% for the test, unlike US.

Lack of competition

The ROL only specifies a credit value and not whether the student attained a merit or excellence level. Some students saw this as an invitation to complete only the minimum requirements for competence. Others felt that the resulting lack of competition for higher grades detracted from their enjoyment of the subject. In contrast, an aspect of assessment for SFC that received favourable comment from students was the motivating power of ranking and competition for grades. One student commented:

> SFC grades motivate me to get a better result because I can try to improve my ranking. You can't improve on getting credit for a Unit Standard, either you get it or you don't.

Disproportionate allocation of credits

Some students expressed concern that the number of credits for each PUS was not proportional to the amount of work required to meet the standard. As one student put it:

> ... some of the Unit Standards do not have enough credit for the amount of work involved in getting the elements whereas in others you can get credit for relatively little work.

They also said that the difficulty and work required to obtain a certain number of credits at Level 2 of the NQF was different for different subjects and yet they all contributed equally towards the National Certificate.

Lack of acceptance by employers and tertiary institutions

There was concern by a group of respondents that the PUS credits were not widely accepted by employers or tertiary institutions. One student explained:

> The Unit Standard credits mean nothing to me. All I worry about is my SFC grades because they will get me into university. I don't think employers understand Unit standards either. They are more concerned about SFC grades that enable them to rank job applicants.

As long as there are two qualifications at the Year 12 level students are going to favour one over the other. One student dismissed the PUS:

> The Unit Standard assessments are just practice for the SFC assessments that really count.

Over-assessment

Another criticism of assessment against the PUS was that students disliked school-based continuous assessment because of its negative impact on student workload. As one student said:

> You've got to work throughout the whole year and there are a lot more tests to do. There are so many assessments and reassessments that it slows us down and this interferes with learning new topics.

Inconsistencies between schools

Consistency between schools which is mentioned above as one of the likes by students was also mentioned as an area of concern by a small number of students. Some felt that success in US assessment depended on the teacher and that this caused variations between schools both in course content and in the number of reassessment activities offered. As one student commented:

> I think it depends very much on the teacher how well you do. Different schools have different standards.

Difficulty of achieving credit

Lower ability students complained that it was too difficult to achieve credits for some PUS and that it was possible to complete the year with only a few credits even though they had completed a substantial number of performance criteria for a number of PUS. This echoes concern raised earlier that the performance criteria may be too specific.

4.4.2 The impact of assessment against the PUS on student learning

Each year teachers who assessed against the PUS were asked to describe the impact of assessment against the PUS on student learning.[27] In 1996 some teachers commented that they found this question difficult to answer:

> ... because students of different abilities responded differently to the Physics Unit Standards.

This observation could not be investigated from the data collected in 1996 but was pursued in the 1997 and 1998 questionnaires by asking teachers about the effect of the PUS on the learning of "More able", "Average" and "Less able" students.[28] The results are summarised in Table 4.16.

Table 4.16: Teachers' perceptions of the impact of assessment against the Level 2 PUS on student learning (1996-1998)

		Percentage of teachers						
		1996 A	1997 B	1997 C	1997 D	1998 B	1998 C	1998 D
5	Very positive	4	10	10	4	20	9	6
4	Positive	35	38	40	31	41	48	31
3	Not sure	44	29	23	15	24	22	18
2	Negative	17	17	25	23	9	17	30
1	Very negative	0	6	2	27	6	4	15
No of respondents		26	48	48	48	54	54	54
Mean response		3.3	3.3	3.3	2.6	3.6	3.4	2.8
Standard deviation		0.8	1.1	1.0	1.3	1.1	1.0	1.2

Key:

A= Students in general B= More able students

C= Average students D= Less able students

[27] 96TQT8e

Table 4.16 illustrates that in 1996, 39% of teachers felt either "Positive" or "Very positive" about the impact of assessment against the PUS on student learning and only a small minority (17%) felt the impact was "Negative". Teachers who felt positively outnumbered teachers who felt negatively byapproximately two to one. A prominent feature of the data was that a large minority (44%) of teachers was "Not sure". The overall patterns for 1997 and 1998 were similar *($F\ (1,300)_{1996/1997/1998} = 2.79$, $p > 0.05$)*.

In 1997 and 1998 a substantial proportion (50%-57%) of teachers considered assessment against the PUS to have had a "Positive" or "Very positive" effect on the learning of "Average" students. Approximately the same proportion of teachers (48%-61%) felt that it had a similar effect on the learning of "More able" students. In contrast to this approximately half (50%-45%) felt that it had a "Negative" or "Very negative" effect on the learning of "Less able" students. A possible explanation for this finding is reflected in the following teacher comment:

> Less able students get discouraged because the level at which the standard is set is too high. They find it difficult to achieve any credit at all. This adversely affects their attitude to physics.

The ANOVA showed that this difference in impact between "More able" and "Average" students on the one hand and "Less able " students on the other hand was statistically significant *($F\ (2, 300)_{\text{Less able/Average/More able}} = 12.71$, $p < 0.0001$)*. The post-hoc Scheffé analysis showed that teachers considered the impact of assessment on students to be similar for "Average " and "More able " students but felt that the impact on "Less able" students *($M = 2.7$, $SD = 1.2$)* was less positive *($p < 0.001$)* than the impact on "Average " students *($M = 3.4$, $SD = 1.0$)* and also less positive *($p < 0.001$)* than the impact on "More able " students *($M = 3.5$, $SD = 1.1$)*.

The conclusion that can be drawn from this is that teachers felt that assessment against the PUS had a positive impact on the learning of able

---

[28] 97TQ20a,b, c, 98TQ20a,b,c

and average students and a slightly negative impact on the learning of less able students. A possible reason for this is that less able students may find it too hard to reach the competency level required for attaining credit for the PUS. Consequently their ROL may show little or no credit. This can be quite disheartening for students and raises the question of the effect of assessment against the PUS on the motivation of students. Under SFC assessment these students would get a grade and some recognition of their achievement.

4.4.3 The impact of assessing against the PUS on student motivation

Each year of the study, teachers were asked for their opinion on how assessing against the PUS affected student motivation in their physics course.[29] The responses are summarised in Table 4.17.

Table 4.17: Teachers' opinions on the impact of assessing against the PUS on student motivation (1996-1998)

		Percentage of teachers		
		1996	1997	1998
5	Very favourably	0	2	4
4	Favourably	28	43	43
3	Not sure	36	32	37
2	Unfavourably	32	21	16
1	Very unfavourably	4	2	0
No of respondents		26	47	51

($\chi^2_{1996/1997/1998} = 4.54$, $df = 4$, $p > 0.05$)[30]

In 1996, only 28% of teachers felt that assessment against the level 2 PUS impacted "Favourably " on student motivation. In 1997 and 1998 approximately 45% of teachers felt this way. Each year about a third of

[29] 96TQT2d, 97TQ16, 98TQ16

teachers (36%-32%-37%) were "Not sure" and a minority (36%-23%-16%) felt that it impacted "Unfavourably" or "Very unfavourably". There were no statistically significant longitudinal differences.

The impact of assessment on the motivation of students was not investigated separately for students of different abilities. Teacher comments however, reflected that the impact differed depending on students' ability.

On the 1998 questionnaire teachers commented that:

> The Physics Unit Standards did not act as a motivator for above average students because they did not recognise or reward excellence. They had the biggest impact on the motivation of average students. They work harder to complete a Unit Standard if they have already got a couple of elements and see a chance to improve. The opportunity to resubmit and improve their results is an excellent motivator. Furthermore the awareness of reassessment has focused them on the learning objectives.

It was felt that assessment against PUS had a negative impact on the motivation of less able students. Teacher D commented:

> Low achievers are pessimistic because they think that they cannot do them so they do not really try. Consequently they don't achieve and because they have no record of success they have totally given up.

Dual assessment was seen to have a negative impact on student motivation. One teacher said that students:

> .... would be better motivated if we stopped having separate SFC assessments and based their SFC grade completely on their achievement in assessments for the Physics Unit Standards.

[30] Categories 1-2 and 4-5 were combined.

4.4.4 Impact of the type of assessment on overall student enjoyment of the Year 12 physics course

Each year of the study students were asked the how enjoyable they found their study of physics that year. [31] The aim of the question was to determine whether the level of student enjoyment of their Year 12 physics course was affected by how they were assessed. Table 4.18 provides a summary of student responses broken down by method of assessment.

Table 4.18: Comparison of students' level of enjoyment of their Year 12 physics course for students assessed against the Level 2 PUS and students assessed for SFC only (1996-1998)

		Percentage of students					
		PUS			SFC		
		1996	1997	1998	1996	1997	1998
5	Very enjoyable	6	5	10	3	7	8
4	Enjoyable	48	47	54	45	38	45
3	Not sure	32	28	24	31	44	33
2	Unenjoyable	11	16	10	18	9	11
1	Very unenjoyable	3	4	2	3	2	3
No of respondents		377	234	163	472	273	338
Mean		3.4	3.3	3.8	3.3	3.4	3.5
Standard deviation		0.9	0.9	2.9	0.9	0.8	0.9

Each year a majority (54%-52%-64%) of students, who were assessed against the Level 2 PUS, found the course "Enjoyable" or "Very enjoyable". A minority (32%-28%-24%) were "Not sure" and an even smaller minority (13%-20%-12%) found the course "Unenjoyable" or "Very unenjoyable". There were statistically significant longitudinal differences in students' opinion *(F (2, 765)*

[20] 96SQT1, 96SQN1, 97SQ2, 98SQ2

$_{1996/1997/1998} = 5.8$, $p < 0.01$). The post-hoc scheffé analysis indicated that students' enjoyment of the course was significantly higher in 1998 than in 1996 ($p < 0.01$) and 1997 ($p < 0.01$).

In contrast, each year approximately half (48%-45%-53%) of students who were assessed for SFC felt that the course was "Enjoyable" or "Very enjoyable". Each year about a third 31%-44%- 33%) was "Not sure" and a smaller proportion (21%-11%-13%) felt it was "Unenjoyable" or "Very unenjoyable". These changes were also statistically significant ($F(2, 1074)_{1996/1997/1998} = 6.45$, $p < 0.01$). The post-hoc Scheffé analysis indicated that students' enjoyment of the course increased significantly from 1996 to 1998 ($p < 0.01$) but was similar in 1997 and 1998.

There was a significant difference in the level of course enjoyment between students who were assessed against the Level 2 PUS and students who were not assessed against the Level 2 PUS ($F(1, 1851)_{PUS/SFC} = 6.19$, $p < 0.01$). The students who were assessed against the Level 2 PUS ($M = 3.5$, $SD = 0.9$) enjoyed their Year 12 physics course slightly more than the students who were assessed for SFC only ($M = 3.4$, $SD = 1.7$).

Students' written replies to this question were divided into two categories. The first included general aspects that affected students' enjoyment of their course. The second included reasons that specifically referred to the type of assessment.

General aspects that students enjoyed[32] included the fascinating nature of the subject, an interesting teacher, a variety of teaching strategies, and practical work and investigative work. Aspects they did not enjoy included the difficulty and mathematical nature of the subject, poor quality teaching and boring teaching strategies.

[32] 96SQT2, 96SQN2, 97SQ3

The following comments related specifically to assessment. Students who were assessed for both the PUS and for SFC did not like the large amount of assessment that resulted from dual assessment. One student said:

> I found it hard juggling SFC and US assessments. More time was spent on assessment than learning and this was a waste of my time.

A reason for not enjoying the course given by PUS students who found it difficult to achieve credit was:

> The Physics Unit Standards are too hard because hardly anyone passes and so you don't want to keep trying, it is very discouraging.

4.4.5 Teacher enthusiasm

Each year of the study, teachers who had assessed against the Level 2 PUS were asked how enthusiastic they felt about this new form of assessment.[33] Their responses are displayed in Table 4.19.

Table 4.19: Teachers' level of enthusiasm towards assessment against the Level 2 PUS (1996-1998)

		Percentage of teachers		
		1996	1997	1998
5	Very enthusiastic	13	11	5
4	Enthusiastic	43	27	43
3	Not sure	23	21	23
2	Unenthusiastic	17	32	17
1	Very unenthusiastic	4	9	12
No of respondents		24	56	58

($\chi^2_{1996/1997/1998} = 8.25$, $p > 0.05$)

[33] 96TQT8f, 96TQN6e, 97TQ3, 98TQ3

Table 4.19 indicates that each year approximately half (56%-38%-48%) of the teachers surveyed felt either "Enthusiastic" or "Very enthusiastic" about assessing against the Level 2 PUS. Each year about a quarter (23%-21%-23%) of teachers were "Not sure" and a sizeable minority (21%-41%-29%) were either "Unenthusiastic" or "Very unenthusiastic". There were no statistically significant longitudinal effects.

4.4.6 The impact of assessment against the PUS on classroom teaching

Each year of the study teachers who assessed against the PUS were asked to describe the effect of this on their classroom teaching.[34] The results are summarised in Table 4.20.

Table 4.20: Teachers' opinions on the impact of assessing against the PUS on classroom teaching (1996-1998)

		Percentage of teachers		
		1996	1997	1998
5	Very positive	9	4	12
4	Positive	30	38	38
3	Not sure	35	28	23
2	Negative	26	24	25
1	Very negative	0	6	2
No of respondents		24	50	52

($\chi^2_{1996/1997/1998}$ = 3.66, df = 6, p > 0.05)[35]

Each year a reasonably large proportion (39%-42%-50%) of teachers who had assessed against the Level 2 PUS felt that this had a "Positive" or "Very positive" impact on their classroom teaching. Approximately 30% (26%-30%-27%) felt that it had a "Negative " or "Very negative" impact on their teaching

[34] 96TQT8d, 97TQ19, 98TQ19

and a similar sized group (35%-28%-23%) were "Not sure". There were no statistically significant longitudinal changes.

In addition to the global quantitative description reported in Table 4.20, teachers were asked to describe advantages and disadvantages associated with assessment against the PUS and the impact on their classroom practice. Table 4.21 provides a summary of the full range of positive and negative impacts identified by teachers. The impacts are not presented in order of frequency of responses.

Table 4.21: Impact of assessment against the Level 2 PUS on classroom teaching

Positive impacts	Negative impacts
• Clearly specified learning outcomes • Professional development • Teacher and programme evaluation • Diagnostic assessment • Motivation of average students • Validation of school-based courses • Seamlessness of the NQF • Course design flexibility	• Excessive workload • Negative impact on teacher morale • Excessive administrative requirements • Pressure on teaching time • Atomisation of learning outcomes • Domination of school programmes • Political and industrial uncertainty

The following sections describe the comments listed in the table in greater detail. While the issues identified by teachers were similar for each year of the study, the comments tended to become more positive over the period of the study. This could in part be due to the fact that teachers who felt that the impact was negative may have withdrawn from assessment against the Level 2 PUS after their initial trial of the system in 1996.

[35] Categories 1 and 2 were combined.

a) Factors that impacted positively

Each year, teachers identified the following aspects of assessing against the PUS that impacted positively on their classroom practice:

Clearly specified learning outcomes

Teachers valued having clear objectives to assess against and to identify what a student can and cannot do. They liked the close links between the PUS and the physics curriculum. One teacher said:

> The performance criteria set out clear expectations of what content needs to be covered. Assessment against the PUS forced me to take a new look at the new curriculum and implement it. I expect mastery and try harder to achieve it.

Professional development of teachers

Teachers felt that the implementation of the PUS assisted their professional development through the three-day training programme, the interaction with the moderator and the available resource material. This was especially reflected in the comments made by sole charge teachers and teachers who were geographically isolated.

Teacher and programme evaluation

Assessment against the PUS provided the teacher with a class profile of results that could be used diagnostically to evaluate programmes. Teachers found this useful. One teacher commented:

> I get better feedback on my teaching. This makes it easier to evaluate my teaching strategies. It has highlighted weaknesses in my teaching and made me think more deeply about assessment tasks. It helps me to identify aspects that should be improved which were previously masked by global percentage student results in tests for SFC.

Another teacher commented:

> My teaching has become more responsive to student needs. I feel I am teaching students and not just content with pre-determined

> assessment dates. I now try to determine when students are ready to be assessed and this reduces the amount of reassessment. Overall assessing against US has impacted positively on my teaching.

Diagnostic assessment

Assessment against the PUS is diagnostic and can identify specific areas of the course with which an individual student is having difficulty. It is also ipsative because it helps individual students by providing guidelines for future learning. One teacher commented:

> The Physics Unit Standards help me to focus my teaching on areas which students have difficulty with. It gives both the students and me a profile of what they can and cannot do. These areas can then be targeted for revision. This is far more useful than a blanket percentage mark at the end of a unit.

Motivation of average students

Assessment against the PUS helped teachers to motivate average students by providing the opportunity to achieve a small unit rather than being overwhelmed by the requirement to get a pass in the course as a whole. It provided students with the possibility of a fresh start for a new unit rather than the cumulative progression towards failure. Teachers commented that this form of assessment does not have the built in failure rate associated with norm-referenced assessment and "eliminates the half right is good enough syndrome". Teachers commented however that assessment against the PUS did not act as a motivator for above average students who wanted recognition of excellence and the below average students who often found it difficult to achieve any credit.

Validation of school-based courses

The old school based SFC courses like Electronics never gave students a nationally recognised qualification. One teacher said:

> I assess against the Level 2 Electronics US as part of my physics course and students can get credit on the Framework. Last year under

> SFC, my electronics work was not recognised separately as counting towards a National Certificate.

Seamlessness of the NQF

Teachers liked the seamlessness between school courses and further tertiary study. One teacher commented that the:

> Physics Unit Standards provide a link between secondary school and further study. The units with which students are credited at school provide the first building blocks of qualifications that can be completed in tertiary study or the workplace.

Course design flexibility

A group of teachers referred to the fact that the Framework allowed greater course flexibility. Some assessed against Level 1, 2 and 3 in a Year 12 course, whereas others had introduced electro-technology units in their programme. One teacher commented:

> The Framework has revitalised my teaching by enabling me to pursue some areas of long-term interest that the narrow SFC prescription did not allow.

a) Factors that impacted negatively

Each year, teachers identified the following aspects of assessment against the PUS that impacted negatively on their teaching:

Excessive workload

Teachers saw the workload associated with assessing against the PUS as its major disadvantage. One teacher summed this up as:

> ... a massive workload for teachers as they grappled with task design, moderation requirements, marking, record keeping and providing reassessment opportunities. It put us under a lot of stress.

Negative impact on teacher morale

Teachers felt that the increase in the workload associated with the implementation of the PUS at a time when other demands such as the implementation of the new curriculum harmed their morale. One teacher lamented:

> I'm trying to be positive and intend offering the Physics Unit Standards again next year but if my workload continues like this next year I will not be keen to continue. I am losing the energy I used to have. I have never had to put so much time into my job and I have been teaching for 30 years. I am submerged in a paper avalanche and am weeks behind in my assessments. I can't imagine doing a Year 13 programme of Unit Standards as well. I am feeling that all I want to do is get out. Maybe I am past it.

Excessive administrative requirements

Teachers complained about the excessive administrative requirements associated with assessment against the PUS. One teacher commented:

> I was very enthusiastic until I actually trialled the standards and became disillusioned with the administrative requirements of moderation, recording, portfolio keeping, filing and reporting. The paperwork requirements and the plethora of moderation forms are mind boggling.

To alleviate this situation, teachers requested the provision of a time allowance, administrative support and increased resources such as item banks of pre-moderated activities. The following comment is a typical example of this type of comment:

> I would like to do Unit Standards but don't have the time to write the assessments and get them moderated. Having to write your own assessment activities is the killer. We need a huge supply of useable activities. It would help if I didn't have to reinvent the wheel with each assessment activity. I think all assessment tasks should come from a

> central pre-moderated pool. I am sure there is enough material out there now. Let's share it. This would make Unit Standards great!

Pressure on teaching time

Teachers commented that the large amount of assessment and reassessment diverted time away from teaching and the development of new strategies. One teacher commented:

> I'm unhappy about Unit Standards! In fact I believe it is a waste of my time. I'd rather be in class teaching than worrying about Unit Standards.

Atomisation of learning outcomes

Some teachers felt that assessment against the PUS was too prescriptive and stifled teacher creativity when designing interesting assessments. The following comment illustrates this view:

> The assessment activities are all of a uniform prescribed format that does not enable me to design unusual questions, use creative contexts or write questions that span several content areas. The compartmentalisation of knowledge and the pedantic insistence on every point. The whole approach is overly legalistic. It makes a chore out of learning for both the student and the teacher. The marking schemes are very rigid and given the ambiguity of some questions, unfair and inaccurate.

Domination of school programmes

In order to ensure that students reach the required standard, some teachers taught directly to the PUS and felt that this affected the flexibility of their programmes. One teacher complained:

> I have not enjoyed my physics teaching this year. I feel we have reverted to the worst aspects of formula dominated teaching and assessment is driving my teaching programme. I am forced to teach to the US to get student passes. It has restricted the open ended approach. My previously inspiring lessons have become drudgery

> trying to get every student up to 100%, the good students get bored and numbers are dropping. There is a lot of valuable time wasted.

Another teacher said:

> The standards are too prescriptive. I believe I will end up teaching to the Unit Standards to the detriment of teaching physics.

Political and industrial uncertainty

Teachers felt that the PPTA industrial action and the Green paper on senior secondary school assessment created an environment of political uncertainty about the future of the senior secondary school qualification system that made it difficult to commit to implementing a new system in their schools. One HOD cautioned:

> It takes a lot of time and energy to put this new system in place. The teachers in my department are worried that their investment of time and energy may be wasted if the government changes its mind.

4.5 Conclusion

Based on the results presented in this chapter it may be concluded that assessment against the Level 2 PUS had satisfactory curriculum fidelity and high concurrent validity with assessment for SFC. PUS credits were found to be relatively satisfactory for indicating student achievement in Year 12 physics and the process of awarding credit was transparent to students.

A majority of teachers who had assessed against the PUS felt that this did not enable recognition of excellence and wanted criteria for the achievement of excellence to be built into the PUS.

Students were unsure about the value of results of assessment against the Level 2 PUS for describing achievement or describing how their performance could be improved and felt that the PUS assessment results were less useful than SFC marks for this purpose.

While the consequential validity of assessment against the PUS was less than that of SFC, there was a steady increase in student satisfaction from 1996 to 1998 with the way they were assessed against the PUS. Students liked the clearly stated and explicit learning outcomes and the diagnostic and formative aspects of assessment against US. This coupled with reassessment was seen as a major advantage because students felt it encouraged thorough learning. Students also liked getting credit at the unit level, the flexible credit distribution, the absence of scaling and the independence of subjects that are a feature of competency-based assessment. The seamless nature of the NQF was seen to be relevant to students who wanted to continue tertiary study at a polytechnic or private training establishment.

Students disliked the all or nothing nature of assessment against the PUS and complained about the pedantic nature of the performance criteria. They felt that assessment against the PUS did not enable differentiation between different levels of achievement and did not allow for the recognition of partial achievement or excellence. While lower ability students found it difficult to achieve credit for the PUS, more able students complained about the lack of competition. Students were worried about the possible lack of acceptance of US by employers and tertiary institutions.

While there were implementation problems in 1996 and 1997, by 1998 nearly half of the teachers surveyed were enthusiastic about the PUS and felt that they had impacted positively on student motivation. This was supported by the finding that students who were assessed against the Level 2 PUS enjoyed their course more than students who were assessed for SFC only. Teachers felt that assessment against the PUS impacted positively on the learning of average and more able students but impacted negatively on the learning of less able students.

Teachers identified a number of aspects of assessment against the PUS that impacted positively on classroom practice. They liked the clearly specified learning outcomes that enabled diagnostic assessment and the evaluation of programmes. Teachers felt that assessment against the PUS helped motivate average students. Teachers felt that the credit structure and seamless nature of the Framework allowed for course design flexibility and the validation of school-designed courses. They also liked the professional development associated with the introduction of this new assessment system.

Teachers complained about excessive workload and administrative demands and the pressure on teaching time. This coupled with political and industrial uncertainty was felt to have a negative impact on teacher morale. Teachers did not like the atomisation of learning outcomes.

Chapter 7 discusses the research findings related to the validity of assessment against the PUS in the context of the research literature, and findings related to reliability and manageability.

Chapter 5

Moderation and the Physics Unit Standards

This chapter discusses the results of the longitudinal investigation into the operation of the physics moderation system at Level 2 of the NQF from 1996-1998. The aim of this investigation was to answer the question:

> ***Does the MAP associated with the PUS achieve an acceptable level of national consistency of assessment and comparability between providers?***

To investigate the effectiveness of various aspects of the moderation system in achieving comparability between schools, moderator and teacher questionnaires were administered annually from 1996-98. In addition, moderator, and end-point assessor judgement agreement trials were carried out to investigate longitudinal trends in moderator and assessor consistency. The results of these investigations are presented in the chronological order in which they occur in the moderation process.

5.1 The physics moderation system

The MAP for the Level 1-4 PUS was implemented in 1996 when 122 providers were allocated to physics moderators. These providers were secondary schools that participated in the 1996 Level 2 PUS trial. The number of providers fluctuated throughout the year as new providers entered the trials and the PPTA imposed a freeze on new developments associated with the NQF.

5.1.1 The moderators

The number of moderators that were active at the start of each year of the study increased from 17 in 1996 through 27 in 1997 to 24 in 1998. Most of the foundation group of moderators continued in their position throughout the period of the study. The length of teaching experience of the 1996 group of moderators ranged from six to 38 years *(M = 18 years, SD =11 years).* They had taught Year 12 physics for an average of 14 years *(SD = 11 years).* This remained similar over the period of the study.

Of all the physics moderators who were employed over the period of the study, 59% had a university qualification which included physics to year three or higher, including two moderators with an honours degree in physics and two with a doctorate in physics. The remainder had stage two physics papers in their degree.

5.1.2 The allocation of providers to moderators

Table 5.1 indicates the total numbers of providers that were hooked on to the physics moderation system at the start of each year. It also shows the mean and range of the allocation of providers to moderators for 1996-98. Figures refer to the start of each year and were subject to fluctuation during the year. The National Moderator and Regional Moderators were allocated fewer providers than Local Moderators because of their higher administrative workload and the extra work associated with check moderation.

Table 5.1: Total number of providers (N) and range and mean number of providers allocated to moderators (1996-1998)

	1996 N = 122		1997 N = 117		1998 N = 150	
	Range	Mean	Range	Mean	Range	Mean
National Moderator	NA	4	NA	6	NA	6
Regional Moderators	2-7	5.5	1-3	1.8	5-8	5.5
Local Moderators	7-11	8.0	1-7	4.7	2-14	6.4

In 1996 and 1997, the main consideration in the allocation of providers to moderators was the manageability of the workload associated with the number of providers. This meant that the main criterion was the number of providers per moderator. Geographical grouping of providers and moderators was a secondary consideration since the moderation system operated predominantly by post. Consequently, the Local Moderator was not necessarily resident in the same location as their provider schools. For example, a Local Moderator in Christchurch moderated the assessment of Southland schools. This system adversely affected interaction between moderators and providers and made it difficult for Local Moderators to organise meetings with their providers.[1]

In 1998, a different system of allocating providers to moderators was used to improve communication. This new approach consisted of grouping providers into geographical regions, and assigning a Local Moderator to a geographical region. A negative consequence of this was that since there were more providers in some regions than others, the workload of moderation was unevenly spread. This explains why the range of providers per moderator

[1] See § 5.3.7 for a discussion of local provider meetings.

increased in 1998. The moderator survey conducted early in 1998 indicated that 95% of moderators were satisfied with the geographical spread and location of their cluster of providers, but not with the number allocated. As a result, some reallocations of providers to moderators occurred later in 1998. A positive aspect of the new system was that it made it easier to hold the annual voluntary meetings of Local Moderators with providers.

5.2 Training of the moderators

Moderators were trained at the commencement of each year of the study to accommodate the increase in the number of providers or to replace moderators who had resigned. The training programme included cross-curricular generic sessions on administrative and procedural requirements, and specific training in the moderation of physics assessment activities and schedules. The initial three-day training programme was reduced to two days in 1997 and 1998. The training of moderators is an important first step in the process of establishing consistency between moderators. The following research question addressed this aspect of the training:

How effective was the moderator training in orienting moderators to the MAP?

To investigate this, each year of the study, newly trained moderators were asked the questions:

- ***How successful was the moderator training you received in enabling you to develop a clear view of your role as a Local or Regional Moderator?***[2]
- ***How successful was the moderator training you received in helping you to develop an understanding of the physics MAP?***[3]

[2] 96MQ5, 97MQ17a, 98MQ26a
[3] 96MQ6, 97MQ17b, 98MQ26b

- ***How successful was the moderator training you received in helping you to develop an understanding of how to carry out moderation of assessment activities and schedules?***[4]

Since only a few new moderators were trained in 1997 and 1998, the data from the three years were combined into a single data set and longitudinal trends in the answers to these questions were not investigated. A summary of moderators' opinions about the effectiveness of aspects of the training programmes is reported in Table 5.2.

Table 5.2: Moderators' views on how successful the moderator training programme was in helping them to develop an understanding of the moderator role, the physics MAP and the moderation of assessment activities and schedules (1996-1998)

		Percentage of moderators		
		The moderator role	The physics MAP	Moderation of assessment activities and schedules
5	Very successful	16	16	22
4	Successful	75	69	69
3	Not sure	9	12	6
2	Unsuccessful	0	3	3
1	Very unsuccessful	0	0	0
Number of respondents		32	32	32

Table 5.2 indicates that the moderator-training programme was highly effective. Over the three years of training, a large majority (91%-85%-91%) of moderators felt that the training programme was either "Successful " or "Very successful" in helping them to develop an understanding of their moderator

[4] 96MQ7, 97MQ17c, 98MQ26c

role, the physics MAP and the moderation of assessment activities and schedules. A small minority (9%-12%-6%) was "Not sure" and only one moderator felt that the training was unsuccessful.

In 1996 none of the prospective moderators had used the PUS as a provider and they were generally unfamiliar with the PUS. To address this the moderator training programme provided practice in the writing of assessment activities and schedules. One of the moderators commented:

> The task of writing activities was valuable. It helped me both as a moderator and as a provider. I appreciated the opportunity to have my efforts critiqued by such a supportive group. I also learnt how to write an assessment schedule with unambiguous judgement statements.

The MAP was well explained by NZQA staff. One moderator remarked:

> The Moderation Action Plan is fairly well set out and easy to follow and the moderation process was clearly described during the training. The step by step approach allowed plenty of opportunities to ask questions.

Comments showed that despite this, some moderators still felt unsure about the MAP after the training. This is not surprising given that there were 15 moderation pro-formas and a number of procedures with which they needed to become familiar.

Several moderators commented that the training gave them a good overview of the administrative requirements of the moderation process, but provided only a brief introduction to the practice of moderation. In 1996 one moderator said:

> The training was a good introduction but further learning occurred on the job and was supplemented by newsletters, the *Assessment Guide* and material from the cluster groups, all of which increased my understanding of my role as moderator.

Despite the success of the training programme, moderators identified various ways in which the moderator training could be improved. In 1997 several moderators commented that there was insufficient time allocated to the training and that the training time should be increased by a day. A Local Moderator commented:

> The training was too crammed and insufficient time was available for following up on how much was understood. We could have done with an extra day.

In 1996 moderators commented that the training could be more successful if specific problems were looked at, and actual activities and student work submitted by providers were used during the training. A typical moderator comment that illustrated this was:

> More time needed to be spent on actually moderating activities. If there had been more time available we could have moderated more activities from a range of standards and compared and discussed notes.

This suggestion was followed up in 1997 and 1998. The group of moderators to be trained was smaller and had a higher level of prior knowledge due to their experience as providers. The training could focus on the techniques and skills of moderation as opposed to the writing process. The training was changed to include more moderation of actual assessment activities that were submitted by providers and included aspects on which moderator opinion was potentially likely to vary.

One moderator commented:

> My use of the Physics Unit Standards as a provider last year gave me knowledge of how to write activities and write and apply assessment schedules. What I wanted was practice in moderating activities and the training provided this.

A general view expressed each year, was that the moderator training provided them with the basic skills of moderating assessment activities, but that:

> The on-the-job experience is necessary to round off the moderator training.

5.3 The moderation system and comparability between providers

Each year of the study, the physics moderators and teachers were asked about the nature of the internal moderation process used at their school as well as a series of questions to address the research question:

> ***How satisfactory are each of the following aspects of moderation in achieving comparability between providers:***

- ***Internal moderation procedures***
- ***moderation of the assessment plan***
- ***moderation of assessment activities***
- ***verification of assessor judgements***
- ***check moderation***
- ***communication within the moderation system***
- ***moderator and provider meetings?***

The following sections discuss each of these aspects of moderation of assessment against the Level 2 PUS in detail. If a clear majority of respondents each year found the aspect "Satisfactory" or "Very satisfactory" then the aspect was deemed to be make an effective contribution to achieving comparability between schools.

5.3.1 Internal moderation procedures

Internal moderation procedures are outside the jurisdiction of the physics MAP. It is however a requirement of accreditation as a provider that satisfactory internal moderation procedures are in place. Comparability between different teachers in the same school is a pre-condition for achieving comparability between schools. In 1996 teachers were asked to describe the procedures used within their school for ensuring that the assessment activities and assessor judgements of teachers for different Year 12 physics classes were of a comparable standard. The answers to this question were analysed separately for schools assessing against the PUS and schools assessing for SFC only. A high proportion of schools that assessed against the PUS practised dual assessment for the PUS and SFC. Apart from statistical scaling for SFC, similar school-based moderation procedures applied to the two modes of assessment. The various methods of internal moderation employed by schools did not change over the period of the study.

Teachers at schools with only one Year 12 physics class did not need any form of internal moderation. Likewise, schools that had only one physics teacher who taught more than one physics class and used the same assessment activities for both classes did not need any internal moderation

The most common form of school-based inter-class moderation for both schools that assessed against the PUS and those who assessed for SFC was to assess all the classes concurrently using common assessment tasks. There were two basic variants on how marking was carried out.

In some schools one teacher marked all of the students' scripts for all classes. To spread the workload, teachers took turns at this throughout the year. This approach aided consistency of end-point assessor judgements but did not take into account that some teachers are habitually easy or hard markers or may have fluctuated in consistency over the marking period.

In other schools marking was carried out by individual class teachers. In addition, teachers marked a sample of their colleagues' scripts and discussed differences. For this option some schools had department meetings to discuss the assessment schedule beforehand and to mark guinea pig scripts to ensure consistency of assessor judgements. In some cases only borderline scripts were discussed.

Schools that assessed for SFC only used all of the above methods of internal moderation but, in addition, used statistical procedures to scale the class test mark distributions from each different class to the mark distribution that class achieved in common exams.

5.3.2 Moderation of the assessment plan

At the commencement of each year, providers were required to submit an assessment plan to their Local Moderator. The plan outlined all the PUS that the provider planned to assess against and the approximate assessment dates. This enabled moderators to select which PUS assessments should be moderated each year. This had to take into account which PUS were moderated the previous year, avoid the PUS which the provider assessed using pre-moderated activities and include the PUS which the National Moderator had targeted for moderation each year.

In 1997 and 1998, moderators were surveyed on how effective they thought the moderation of the assessment plan was in achieving comparability between schools.[5] Table 5.3 contains a summary of moderators' responses.

[5] 97MQ3a, 98MQ4a

Table 5.3: Moderators' views on how satisfactory the moderation of the assessment plan was in achieving comparability between schools (1997-1998)

		Percentage of moderators	
		1997	1998
5	Very satisfactory	20	15
4	Satisfactory	40	65
3	Average	25	10
2	Not satisfactory	15	0
1	Very unsatisfactory	0	10
Number of respondents		22	20

Table 5.3 illustrates that in 1997 and 1998 a substantial majority (60%-80%) of moderators felt that the moderation of the assessment plan was "Satisfactory " or "Very satisfactory". Over the same period only a minority (15%-10%) felt it was "Not satisfactory" or "Very unsatisfactory". Longitudinal effects were not investigated because a χ^2 analysis could not be carried out. The reason for this was that data did not meet the requirement that the expected values should be greater than 5 for at least 80% of the cells (Burns 194:178).[6] The changing nature of the comments however suggested an increasing level of moderators' confidence.

A typical reason moderators gave for being satisfied with the moderation of the assessment plan was that it helped teachers plan their assessment programme in advance and encouraged them to write their assessment activities ahead of time to meet the moderation deadlines. One moderator felt that:

[6] The χ^2 analyses throughout this chapter were carried out on the frequencies of responses not the percentage data displayed in the tables.

> ... this led to better planned and more valid assessment activities that are of higher quality and presentation than those produced at the last moment.

Moderators also felt that the targeting of specific PUS for moderation each year helped to achieve comparability. One moderator remarked:

> The moderation of the assessment plan does not directly contribute to comparability. It does make it possible however for moderators to select the same Physics Unit Standard for moderation from different schools. This makes it easier to establish consistency than if a wider range of standards were moderated.

Moderators who were dissatisfied with the moderation of the assessment plan perceived its main aim as administrative rather than contributing towards comparability. The following moderator comment reflects this view:

> Moderation of the assessment plans does not directly contribute to achieving comparability between schools but is basically an administrative task to help plan moderator workload.

Based on these findings it can be concluded that the moderation of assessment plans made an effective contribution to the physics MAP in its aim to achieve comparability between schools. It was also seen as a tool to manage moderator workload.

5.3.3 Moderation of assessment activities

The physics MAP required that each year the assessment activities for 20% of the Level 2 PUS in a provider's programme were moderated. The moderation of assessment activities provided a check on the curriculum fidelity of the providers' assessment programmes and aimed to contribute to comparability between providers. Each year of the study moderators were asked how satisfactory they found the moderation of assessment activities in achieving this aim.[7] Table 5.4 summarises their responses.

[7] 96MQ8a, 97MQ3b and 98MQ4b

Table 5.4: Moderators' views on how satisfactory the moderation of assessment activities was in achieving comparability between schools (1996-1998)

		Percentage of moderators		
		1996	1997	1998
5	Very satisfactory	0	29	28
4	Satisfactory	54	46	62
3	Average	31	17	10
2	Not satisfactory	15	8	0
1	Very unsatisfactory	0	0	0
Number of respondents		13	24	21

($\chi^2_{1996\text{-}1998} = 8.79$, $df = 6$, $p > 0.05$)[8]

Table 5.4 illustrates that each year of the study, a clear majority (54%-75%-90%) of the moderators surveyed thought that the moderation of assessment activities made a "Satisfactory" or "Very satisfactory" contribution to achieving comparability between schools. Over the same period, a minority (31%-17%-10%) felt that it made an "Average" contribution and no moderator felt that it was "Very unsatisfactory". There were no statistically significant changes in moderator opinions over this period.

The main reason for moderator confidence that moderation of assessment activities was achieving comparability between providers is reflected by the following comment made by a moderator in the 1996 questionnaire:

> I am in a position to see assessment activities for the same Unit Standard from different schools and the comparability between providers improved as the year progressed and teachers gained more experience.

[8] Categories 1 and 2 were combined to meet the cell size requirement for χ^2 analysis.

Moderators identified several factors that contributed to comparability. They found that the *Assessment Guide: Physics* (1996c), the additional packs of physics assessment activities (1997) and the CD of Level 2 activities were useful in helping providers and moderators interpret the PUS, particularly in cases where the standards appeared ambiguous. If a provider had difficulty in producing assessment activities of suitable quality, they were frequently referred to the assessment guide or CD for suitable exemplars. Several moderators commented that providers were submitting assessment activities from the *Assessment Guide: Physics* for moderation and that this helped in establishing comparability between schools.

In 1996, a moderator who felt that the moderation of assessment activities was "Not satisfactory", reasoned that:

> It is too early to comment on the extent to which the moderation process is achieving comparability between schools. At this stage, the assessment activities generated by my providers show that there is considerable variation between schools. Hopefully we should achieve better consistency as providers and moderators become more experienced.

In addition moderators argued that since only 20% of a provider's assessment programme was moderated each year moderators were unsure of the quality of the remaining 80%. This was reflected in the comment:

> It is difficult to tell what consistency there is between assessment activities from different schools for US that are not moderated each year.

This should improve over time since the programme of PUS that was targeted for moderation each year aimed to cover all of the PUS in a 4-5 year cycle.

In 1998, comments tended to be more positive and none of the moderators felt that the moderation of assessment activities was unsatisfactory, although 10% were still "Not sure".

Each year a parallel question about the effectiveness of the moderation of assessment activities was asked of teachers who had assessed against the Level 2 PUS.[9] Table 5.5 contains a summary of teacher opinion about the effectiveness of the moderation of assessment activities over the period of the study.

Table 5.5: Teachers' opinions on how satisfactory the moderation of assessment activities was in achieving comparability between schools (1996-1998)

		Percentage of teachers		
		1996	1997	1998
5	Very satisfactory	22	13	2
4	Satisfactory	43	46	45
3	Not sure	13	30	44
2	Unsatisfactory	22	2	9
1	Very unsatisfactory	0	9	0
No of respondents		26	46	55

($\chi^2_{1996\text{-}1998} = 16.99$, $df = 6$, $p < 0.01$)[10]

Table 5.5 shows that in 1996, a majority (65%) of the sample of trial-school teachers felt that the moderation of assessment activities was either "Satisfactory" or "Very satisfactory". This dropped to 59% in 1997 and 47% 1998. This decrease was accompanied by an increase in the percentage of teachers who were "Not sure" from 13% in 1996 through 30% in 1997 to 44% by 1998. This suggests that while teachers did not find the moderation of assessment activities unsatisfactory, they were increasingly unsure of whether it was achieving comparability between schools. This trend was statistically significant and is reflected in the changes in teachers' comments over this period.

[9] 96TQT7c1, 96TQN5c1, 97TQ17a, 98TQ17a
[10] Categories 1 and 2 were combined.

In 1996, teachers commented that they found the moderation of assessment activities useful because it helped to define the standard to write assessment activities to and gave them feedback on their assessment programme. A sole physics teacher at a school remarked:

> I find it good to have someone proof-read and check my work and give me some feedback. As the only physics teacher at this school, I would feel very isolated if it weren't for the contact with my moderator. The moderation process has a strong professional development aspect and should help to achieve a similar standard across schools.

A possible reason for the increase in teacher uncertainty was reflected in the 1997 responses. Some teachers quoted hearsay evidence that some moderators accepted assessment activities that had been previously rejected by another moderator. A possible reason for this is that the language of version one of the PUS was ambiguous in places and meant that different interpretations each meeting the requirements of the PUS were possible. An example of this was the 'extended investigation', which some schools completed in a few periods and others spread over several weeks. This PUS was chosen deliberately for moderation in the first year in order to try and establish a consistent interpretation. The revision of the standards has tightened up the language used and the revised assessment guide now contains activities that were trialled in schools as guides for the interpretation of the PUS.

Comments made in 1998 reflected that the emphasis for teachers in having assessment activities moderated shifted from the professional development aspect of receiving feedback on their own assessment activities to concern about comparability between schools. One teacher commented:

> I have had no information about inter-school comparability but get the feeling that moderators for different schools have different ideas and standards. Some schools may design assessment activities using a

context in which it is more difficult to show competency than contexts used at other schools.

Moderators felt that the moderation of assessment activities made an effective contribution to achieving comparability between schools and in 1996 and 1997 teachers agreed with this view. In 1998 however only 47% of teachers agreed. There is clearly a disparity in the trends between teachers' and moderators' views. Moderators appeared to be increasingly confident that comparability was being achieved whereas teachers were increasingly unsure. Since moderators were in a position to see assessment activities for the same Physics Unit Standard from a variety of schools they were in a better position to judge. This coupled with the longitudinal decrease in the percentage of teachers who found the moderation unsatisfactory, probably means that satisfactory comparability was achieved. The moderator agreement trials reported later in this chapter provide additional data to support this.

5.3.4 Verification of assessor judgements

The MAP required that for each assessment activity that was moderated, providers had to submit six items of marked student work for verification of assessor judgements. The verification of the marking provided moderators with an insight into the consistency of end-point assessor judgements for their allocated providers. Since the assessment schedules had been moderated beforehand, the judgement statements should precisely specify the requirements for student answers to meet the performance requirements, and therefore result in high consistency in marking. Whether this was occurring was investigated each year, by asking moderators how satisfactory they found the moderation of assessor judgements in achieving comparability between schools.[11] Table 5.6 summarises their responses.

[11] 96MQ8b, 97MQ3c, 98MQ4c

Table 5.6: Moderators' views on how satisfactory the moderation of assessor judgements was in achieving comparability between schools (1996-1998)

		Percentage of moderators		
		1996	1997	1998
5	Very satisfactory	8	27	38
4	Satisfactory	53	59	47
3	Average	31	9	10
2	Not satisfactory	8	5	5
1	Very unsatisfactory	0	0	0
Number of respondents		13	22	21

($\chi^2_{1996\text{-}1998} = 3.80$, $df = 2$, $p > 0.05$)[12]

Table 5.6 shows that moderators were generally confident that the moderation of assessor judgements was effective. Over the three years of the study a definite majority (61%-85%-85%) felt that it was either "Satisfactory" or "Very satisfactory, a minority (31%-9%-10%) were "Not sure" and only a small percentage (8%-5%-5%) considered it to be "Not satisfactory". There were no statistically significant longitudinal differences in this pattern of responses.

Comments made by moderators on the 1996 questionnaire showed that initially there was doubt about the consistency of end-point assessor judgements between providers. They were:

> Surprised at how strict some and how lenient other teachers were.
> There was a tendency by some providers to side-track some judgements and not to be firm when making decisions.

The comments made by moderators in subsequent years tended to be more confident. In 1998 a typical moderator comment was:

[12] Categories 1 and 2 were combined

> The teachers in my allocation were generally very accurate in their marking and I seldom pick up any mistakes. I feel that the number of items of student work that need to be submitted should be reduced.

Since the assessment schedules were moderated before the teachers used them, the judgement statements were generally tight. As a result marking tended to be accurate. This information was relayed to NZQA in the annual moderator reports. For this reason the number of items of student work which needed to be submitted to the moderator for verification of assessor judgements was reduced from six to four in 1998.

Each year of the study, teachers who had assessed against the Level 2 PUS were asked a similar question about the moderation of assessor judgements.[13] Table 5.7 contains a summary of their responses.

Table 5.7: Teachers' opinions on how satisfactory the moderation of assessor judgements was in achieving comparability between schools (1996-1998)

		Percentage of teachers		
		1996	1997	1998
5	Very satisfactory	27	16	2
4	Satisfactory	45	49	61
3	Not sure	18	31	33
2	Unsatisfactory	9	0	2
1	Very unsatisfactory	0	4	2
No of respondents		25	45	51

($\chi^2_{1996\text{-}1998} = 11.30$, $df = 4$, $p < 0.05$)[14]

Teachers agreed with moderators that the moderation of assessor judgements was a satisfactory component of the MAP. Table 5.7 illustrates that over the three years of the study approximately two-thirds (72%-65%-

[13] 96TQT7c2, 96TQN5c2, 97TQ17b, 08TQ17b

[14] Categories 1-3 were combined.

63%) of teachers felt that it made a "Satisfactory" or "Very satisfactory" contribution to achieving comparability between schools. The percentage of teachers who were "Not sure" increased from about a fifth (18%) in 1996 to about a third (31%-33%) in 1997 and 1998. Since the percentage of teachers who felt that the moderation of assessor judgements was "Unsatisfactory " or "Very unsatisfactory" decreased over this period (9%-4%-4%), these changes reflect increasing uncertainty rather than increasing dissatisfaction. This trend was statistically significant and is also reflected in the changing nature of teachers' comments over this period.

In 1996 teachers felt that the verification of assessor judgements was needed:

> To maintain comparability between schools, it is very helpful to have someone check your marking and give you feedback that you are on the right track, it helps your confidence no end.

Another teacher reflected:

> I have no basis for judging comparability. I just have to trust that NZQA knows what it is doing. At least with SFC we have a basis for comparison with other schools.

The reason for the increasing uncertainty about the contribution of the moderation of assessor judgements towards comparability between schools, was reflected in the following comment made in 1997:

> The marking of student work was only checked for the 25% of the PUS offered in a school's programme which are moderated and there is no check on whether teachers judge their students more leniently for the remaining units.

In 1998 one teacher remarked that:

> Once the activity is moderated there does not seem to be a need to verify the marking. Surely we should trust teachers to be professionals.

Each year both moderators and teachers felt that the moderation of assessor judgement made an effective contribution to achieving comparability between schools. There was a disparity in the longitudinal trends between teachers' views and moderators' views on the contribution made by the moderation of assessor judgements towards achieving comparability between schools. Moderators appeared increasingly confident although this was not statistically significant whereas teachers were increasingly uncertain. Moderators were probably in a better position to monitor the trend because they verified the marking from a range of different teachers over a three-year period. The end-point assessor judgement agreement trials reported later in the chapter provide a more quantitative answer to this question.

5.3.5 Check moderation

Check moderation is the process whereby the Regional Moderator checks a 20% sample of the moderation carried out by the Local Moderators and the National Moderator checks a 20% sample of the check moderation carried out by the Regional Moderators. It aims to monitor consistency between moderators in order to achieve comparability between schools. In 1997 and 1998, moderators were asked how satisfactory they thought the check moderation process was meeting this aim.[15] Table 5.8 summarises their responses.

[15] 97MQ3e and 98MQ4e

Table 5.8: Moderators' views on how satisfactory the check moderation process was in achieving comparability between schools (1997-1998)

		Percentage of moderators	
		1997	1998
5	Very satisfactory	25	16
4	Satisfactory	75	47
3	Average	0	32
2	Not satisfactory	0	5
1	Very unsatisfactory	0	0
Number of respondents		16	19

Table 5.8 shows in 1997 all of the moderators felt that the check moderation process was either "Satisfactory" or "Very satisfactory". In 1998 about two-thirds (63%) felt this way, about a third (32%) felt it was "Average" and one moderator felt it was "Not satisfactory". [16]

The moderators'comments reflect that as moderators became more experienced they did not necessarily become dissatisfied with the process but may have become less reliant on it.

In 1997 moderators felt that the check moderation process was important as a quality control measure. One moderator commented:

> I have no idea what other moderators are doing. There is comparability between the schools I am moderating, but whether this compares to other schools I don't know. It is important for the Regional and National Moderators to see a cross-section of provider assessment activities and pupils' work. Check moderation is an important component in establishing and maintaining the consistency of moderator decisions between moderators and achieving comparability between schools.

[16] χ^2 analysis was not carried out because of insufficient cell size.

In 1998 moderators were more experienced and were possibly more confident about their decisions. Consequently the role of check moderation was seen as less important. One moderator commented that:

> The check moderation process basically confirmed my moderator decisions, and apart from increasing my confidence in my own moderator decisions, I do not depend on it as much as last year.

A regional moderator commented:

> Most of the time my check moderation endorses the decisions made by the Local Moderator. I do feel however that it is important to have a check like this because I do pick up the occasional inconsistent and wrong decisions.

Each year moderators felt that check moderation made an effective contribution to the MAP in its aim to achieve comparability between schools. It may be that the amount of check moderation can be reduced in time but larger scale studies over a number of moderation systems are necessary before taking this step.

5.3.6 Communication within the moderation system

Communication between moderators and between moderators and providers is an important component of establishing and maintaining consistency. For each year of the study moderators were asked how satisfactory the communication with the National Moderator, the Regional Moderator and the teachers in their allocation was in achieving comparability between schools.

a) Communication with the National Moderator

Communication between the National and Local Moderators was mainly confined to the moderation newsletters and the National Moderator's attendance at regional meetings of moderators. In addition the National Moderator helped sort out disagreements between moderators and between moderators and providers. Table 5.9 contains a summary of moderator opinion on the effectiveness of this communication.

Table 5.9: Moderators' views on how satisfactory communication with the National Moderator was in achieving comparability between schools (1996-1998)

		Percentage of moderators		
		1996	1997	1998
5	Very satisfactory	46	60	55
4	Satisfactory	38	36	45
3	Average	15	4	0
2	Not satisfactory	0	0	0
1	Very unsatisfactory	0	0	0
Number of respondents		13	25	20

Table 5.9 shows that each year of the research, a large majority (84%-96%) of moderators felt that communication with the National Moderator was either "Satisfactory" or "Very satisfactory" and none of the moderators felt that communication was "Not satisfactory" nor "Very unsatisfactory". [17]

Comments made by moderators on the 1996 questionnaire[18] indicated that they felt that the Unit Standard interpretations made by the National Moderator which were published in the newsletters to moderators were valuable, particularly in the early stages of establishing the moderation system. Some moderators commented that they wanted the newsletter to be more frequent, perhaps on a monthly basis, whereas others commented that they would like to see a regular newsletter for schools to cover Unit Standard interpretations.

[17] χ^2 analysis not carried out because of insufficient cell size.
[18] 96MQ8d

Moderators felt that the attendance of the National Moderator at all of the regional meetings helped to establish national consistency of moderation and helped to prevent regional differences from occurring. The comments made by moderators on the 1997 and 1998 questionnaires[19] were similar and confirmed that it is necessary to have one person who has an overall perspective of what is happening throughout the country.

b) Communication with the Regional Moderator

Each year of the research, Local Moderators were surveyed on how satisfactory communication with their respective Regional Moderator was in achieving comparability between schools.[20] Table 5.10 contains a summary of their responses.

Table 5.10: Local moderators' views on how satisfactory communication with their Regional Moderator was in achieving comparability between schools (1996-1998)

		Percentage of moderators		
		1996	1997	1998
5	Very satisfactory	31	54	37
4	Satisfactory	54	42	42
3	Average	8	4	16
2	Not satisfactory	8	0	5
1	Very unsatisfactory	0	0	0
Number of respondents		12	24	19

Table 5.10 shows that each year a large majority (85%-96%-79%) of Local Moderators felt that communication with their Regional Moderator was either "Satisfactory " or "Very satisfactory". Only a small minority (8%-4%-16%) felt it

[19] 97MQ3g, 98MQ4g
[20] 96MQ8c, 97MQ3f, 98MQ4f

was "Average". None of the moderators felt it was "Very unsatisfactory" and in 1996 and 1998 only one moderator felt it was "Not satisfactory".[21] The reason for the high level of support may be due to the fact that moderators obviously appreciated the support of their Regional Moderator during a period that saw industrial action, revision of standards and the implementation of a new curriculum.

The purpose of this communication is twofold. The check moderation carried out by the Regional Moderator is a consistency check on the moderation of the moderators in the regional group. The other aspect is to act as a support person for the interpretation of the PUS, dealing with providers and seeking advice on issues, such as, non-compliance of providers. In all of these categories moderators appreciated the support provided by their Regional Moderator and consequently rated this highly in the questionnaire.

One Local Moderator commented:

> I have only had to ask a few questions and she always responds promptly. She is very accessible and supportive. Having access to a Regional Moderator to discuss problems with providers or check on the interpretation of a Unit Standard is essential.

c) Communication between Local Moderators and providers

Each year every provider nominates a contact person for communication with the Local moderator and moderators were asked to describe the effectiveness of the communication with their allocated provider contact persons.[22] Table 5.11 summarises the moderators' opinion on the effect of this communication on the process of achieving comparability between schools.

---

[21] χ^2 analysis not carried out because of insufficient cell sizes.
[22] 96MQ8e, 97MQ3h, 98MQ4h

Table 5.11: Moderators' views on how satisfactory communication with the contact person in provider schools was in achieving comparability between schools (1996-1998)

		Percentage of moderators		
		1996	1997	1998
5	Very satisfactory	8	13	0
4	Satisfactory	38	50	53
3	Average	46	29	31
2	Not satisfactory	8	4	16
1	Very unsatisfactory	0	4	0
Number of respondents		13	24	21

Table 5.11 indicates that each year, a moderate proportion (46%-63%-53%) of the moderators surveyed felt that communication with the provider contact was "Satisfactory " or "Very satisfactory" in contributing to comparability between schools. A sizeable minority (46%-29%-31%) considered it "Average" and only one or two moderators felt it was "Not satisfactory" or "Very unsatisfactory".[23]

Moderators felt that communication between providers and their moderator was essential in establishing and maintaining comparability between schools. One moderator commented:

> Communication between moderators and teachers on the activities and schedules used is vital and that comparability between schools and between moderators needs to be carefully and continuously monitored.

The communication between Local Moderators and teachers was generally satisfactory. One moderator commented:

> The HODs and physics teachers in my provider group are very supportive and communication is generally good. The teachers and I

> value the professional contact and sharing of ideas that has resulted from moderation.

The communication mainly related to assessment activities and schedules but also related to provider queries about the assessment system.

A possible reason some moderators felt that the communication was unsatisfactory is reflected in the following comment made in 1996:

> Communication was good with some schools, hopeless in others. Some providers think they can ignore the moderation process and they rarely initiate contact. Communication with the contact person in provider schools was affected by the NZPPTA Framework freeze.

Moderators felt that this may be due to the heavy workloads experienced by teachers as a result of dual assessment. One moderator lamented:

> ... dual assessment for both SFC and Unit Standards is a killer which leads to an unacceptable workload. For some providers, classes are too large e.g. 38 students per class, they just do not have time for the administration associated with moderation or meeting the deadlines.

One issue that was raised by moderators each year of the study was that some providers had not received training in assessing against the PUS and expected the moderation system to deliver this. The role of the moderator was contractually restricted to the moderation of assessment activities and the verification of assessor judgements but providers expected moderators to have a wider professional development role. This led to conflict in role definition which some moderators had difficulty in dealing with.

Teachers were asked a parallel question on how satisfactory they found communication with their Local Moderator.[24] Table 5.12 contains a summary of their responses.

---

[23] χ^2 analysis not carried out because of insufficient cell sizes.
[24] 96TQT7c3, 97TQ17d, 98TQ17d

Table 5.12: Teachers' opinions on how satisfactory communication with their Local Moderator was in achieving comparability between schools (1996-1998)

		Percentage of teachers		
		1996	1997	1998
5	Very satisfactory	52	20	21
4	Satisfactory	30	57	55
3	Average	0	12	13
2	Unsatisfactory	17	4	6
1	Very unsatisfactory	0	6	6
No of respondents		25	49	53

($\chi^2_{1996\text{-}1998} = 13.37, df = 6, p < 0.05$)[25]

Each year a clear majority of teachers agreed with moderators that the communication between teachers and moderators was effective. Table 5.12 shows that the percentage of teachers who felt that communication was either "Very satisfactory" or "Satisfactory" reduced from 82% in 1996 through 77% in 1997 to 76% by 1998. This was accompanied by an increase in the percentage of teachers who were "Not sure" from 0% in 1996 through 12% in 1997 to 13% by 1998 and a decrease in the percentage of teachers who thought it was "Unsatisfactory" or "Very unsatisfactory" (17%-10%-12%). These trends were statistically significant and indicate that as teachers became more experienced, they were less satisfied with the interaction with their Local Moderator. This may be because they became less dependent on the Local Moderator for guidance.

The moderator role was a source of confusion in the communication between teachers and moderators. Teachers wanted the moderator to act in an advisory capacity but the moderator contract precluded this. One teacher suggested:

[25] Categories 1 and 2 were combined

> NZQA needs to employ subject advisors or work with the Education Advisory Service to provide support to schools which the demands on the moderation system show is lacking.

Problems mentioned by teachers included time delays in returning moderated assessment activities or responding to their queries and, on one occasion, a lack of response from a moderator. One teacher said:

> My moderator has not answered my request. It is now June and I asked in March.

The geographical distance between the moderator and provider made communication difficult when issues which were not easily resolved by fax or post but needed direct communication. This, coupled with the geographical spread of providers, made it difficult to arrange and attend local meetings. One respondent commented that:

> Phone contact is excellent but it is difficult to find a suitable contact time.

Each year moderators felt that communication with the National Moderator and their Regional Moderator was a highly effective aspect of the MAP. Each year a majority of moderators and teachers felt that the communication between providers and moderators was average or better.

5.3.7 Moderator and provider meetings

Another aspect of communication within the physics moderation system was the annual cycle of moderator and provider meetings.

Early each year there was a national meeting attended by the National and Regional Moderators for biology, chemistry, physics and science. This was followed up by a round of regional meetings, between each Regional Moderator and all of the Local Moderators in their regions. In 1996 and 1997, regional meetings were held in Wellington, Christchurch, Auckland and Hamilton. In 1998 there were only three regional meetings. These were held in Christchurch, Wellington and Auckland. These meetings were funded by NZQA. The National Moderator attended all the regional meetings each year to provide an element of consistency in the interpretation of the PUS. The regional meetings were followed up by optional meetings between Local Moderators and their allocated providers. The purpose of all of these meetings was to discuss issues related to the administration of the moderation system and to compare and reach consensus on the moderation of common assessment activities in order to contribute to comparability between schools. In 1998 moderators were surveyed on how satisfactory the moderation meetings were in achieving this purpose.[26] Table 5.13 contains a summary of their responses.

Table 5.13: Moderators' views on how satisfactory local and regional meetings were in achieving comparability between schools (1998)

		Percentage of moderators	
		Regional meetings	Local meetings
5	Very satisfactory	47	0
4	Satisfactory	48	9
3	Average	0	27
2	Not satisfactory	5	37
1	Very unsatisfactory	0	27
No of respondents		11	19

Table 5.13 indicates that in 1998, all but one of the physics moderators felt that the regional meetings were either "Satisfactory" or "Very satisfactory" in achieving comparability between schools. One moderator said:

[26] 98MQ4i,4j

> The positive aspect of the regional meeting is that a great deal of agreement on standards and moderator decisions is reached.

However, 64% of the moderators felt that the local meetings were "Not satisfactory" or "Very unsatisfactory" in achieving comparability between schools. They were also asked whether meetings between Local Moderators and providers were necessary. Moderator opinion was split, with 52% of moderators stating that they felt the meetings were necessary and 48% stating they were not.

While there is strong support for the meetings between moderators, moderators were less sure of the value of the meetings with providers. The following comment made by a Regional Moderator suggests a possible reason for this:

> The meetings with providers were generally unsatisfactory. The meetings were not compulsory and funding was not provided. They had to be held outside school hours, generally on a Saturday. This was not popular with teachers and there was a high rate of absenteeism. Moderators found it difficult to organise these meetings because of the wide geographical spread of their providers.

5.4 The Nationally Prescribed Activity

In 1996 and 1997, the MAP for physics included a Nationally Prescribed Activity (NPA). This was a common assessment activity that was to be administered by all providers within a stated time period.

The research questions related to the NPA were:

- ***What is the role of the NPA in the MAP?***
- ***How satisfactory is the NPA in achieving comparability between providers?***
- ***What should the results of the NPA be used for?***

- ***When should the NPA be held?***
- ***Is the NPA necessary?***

a) What is the role of the NPA in the MAP?

In the 1996 questionnaire moderators commented that NZQA did not collect school results for the NPA at the end of the year and expressed doubt about the purpose of the NPA. This was followed up in 1997 and moderators and teachers were asked what they thought the role of the NPA was. [27]
Moderators mentioned that the role of the NPA was to:

- check the consistency of teacher marking throughout the country and to give feedback to assessors in each school.
- illustrate a national standard for the design of assessment activities and schedules and to illustrate the interpretation of the Unit Standard concerned.
- provide a high quality assessment exemplar which can be used by teachers as a model of good assessment practice.
- moderate the marking of providers and their ability to follow the marking schedule.

Teachers' responses were similar to the moderators but contained the additional reason:

> The NPA provided a check that the level of ability of skills of my students is equivalent to the national standard.

b) How satisfactory is the NPA in achieving comparability between providers?

In 1997, teachers were asked their opinion on how satisfactory the NPA was in achieving comparability between schools.[28] Only 24% of teachers felt that it was "Satisfactory or "Very satisfactory", 48% were "Unsure" and 28% felt it

[27] 97TQ25c
[28] 97TQ17c

was "Unsatisfactory" or "Very unsatisfactory". The following excerpts of comments made by teachers are possible reasons for this:

- it was not compulsory
- this sort of statistical check was part of "the old philosophy"
- the timing of the NPA did not relate to a logical time in a provider's programme
- the NPA should not be a reference test to be taught to as a focus for the year
- the moderation system has enough checks on consistency built into it without adding another one
- often students had already received credit for the Unit Standard assessed by the NPA and were unnecessarily assessed again
- NZQA didn't use the results to check national consistency.

c) What should the results of the NPA be used for?

In 1997 moderators were asked whether the NPA results from each school should be used as a statistical check on the way schools award credit. Only 13% of moderators agreed, 35% were "Not sure" but 52% felt that it should not be used this way. The reason given for this was that:

> There are too many variables and sources of variation for this to be reliable. It would signal a lack of trust in the professionalism of teachers.

Moderators who agreed that the results of the NPA should be used as a statistical check felt:

> ... that it was important for the integrity of the system to have a check like this.

The check recommended by moderators, was a comparison of each provider's distribution of credit for the NPA with the provider's total distribution of credit for the year. This could be used to establish if the proportion of students who reach the standard under the controlled conditions of the NPA was similar to the proportion who reach the standard under the relatively uncontrolled conditions of the 80% of PUS assessments which are not moderated each year.

d) *When should the NPA be held?*

In 1996 the NPA had to be administered within a period specified by NZQA. This led to problems. Each school's order of delivery of the curriculum was different, and therefore, schools were not necessarily ready to administer the NPA at the specified time. Some schools had done the topic on which the NPA was based just prior to the NPA, whereas others had done it months before. Some students were assessed for this PUS for the first time in the NPA. Others had already passed the PUS and didn't see the need to revise. For others it was a reassessment. In 1997 teachers were asked what they thought was the best time to hold the NPA.[29] Teachers' opinions were equally divided between wanting to assess with the NPA at any time during the year (43%) or within a specified period (43%). Only 16% wanted the NPA to be held on a specified day.

In 1997, all of the moderators were in support of having the NPA at a naturally occurring time in a provider's assessment programme and not within an NZQA specified period as was the case in 1996. One moderator commented:

> Schools should be able to use it any time during the year, at a time convenient to them. The philosophy of assessment against the Physics Unit standards is that it should be close to the learning. This would require advance signalling of the Unit Standard number the NPA will assess against and early preparation of the activity by NZQA.

[29] 97TQ25b

e) Is the NPA necessary?

In 1997, moderators were surveyed on whether they thought the NPA was a necessary part of the MAP. The responses indicated that 64% of moderators thought that the NPA was "Necessary" or "Very necessary" to the moderation process, 8% were "Not sure" and 28% felt that it was "Not necessary" or "Very unnecessary". One moderator replied, it is:

> ... needed to nationally check the assessors for consistency and to standardise marking across the country. It is the only time that all students sit the same activity for the same Unit Standard and are marked using the same assessment schedule.

Some moderators expressed the view that:

> It should be compulsory and all providers in the system should have to comply.

In 1997, teachers were asked a similar question.[30] Eighty-two percent of teachers were in favour of the NPA and 18% were not. The reason for teachers' support of the NPA was that it provided a high quality assessment resource. Comments did not however, express support for the quality control aspect of the NPA. One teacher commented:

> The NPA is a good resource but I do not like being told when to run it or have it used to somehow check on the assessment programme in my school.

Teachers in the case-study schools liked the principle of having a national common assessment activity. Teacher A commented:

> The NPA is a useful exercise. The examiner signals the standard and the moderators check that it is consistently marked.

Teacher C added:

> The NPAs are a good assessment resource. I included the 1996 NPA in my school exam. It is hard to achieve this kind of presentation

---

[30] 97TQ25a

> yourself. It raises awareness of how to write assessor judgements and illustrates the format required for the submission of assessment activities to my moderator.

Teachers were not impressed with NZQA's organisation and administration of the NPAs. The NPA was not compulsory and in 1996 teacher D commented:

> The administration of the NPA has loopholes, giving it little purpose. It was a sham, a complete disaster. NZQA did not even know who needed NPAs and had to ask schools. The PPTA freeze did not help. If it is to be truly national it should be compulsory for all schools doing Unit Standards.

Following in-house confidential research by NZQA, the NPA was abolished in 1998. Given the earlier quoted statistics of teacher and moderator support for the NPA, this would at first glance appear to be an unpopular move. In the same year however, NZQA introduced pre-moderated activities. These were different from the NPAs in that they were not compulsory and could be used at any time during the year without having to be submitted for moderation. They were seen to encompass all of the advantages of the NPA without any of the disadvantages. Teacher B said:

> The pre-moderated activities *The physics of Sport* and *Atoms and Radiation* were of a very high standard and raised the professional look of the subject to students. The contexts were interesting and stimulating for students.

In the 1998 questionnaire moderators were asked their opinion on the changes. Ninety percent of the moderators agreed with the decision to abolish the NPA and 95% agreed with the provision of pre-moderated activities. This was popular because the pre-moderated activities did not have to be submitted for pre-assessment moderation and this was seen by teachers as helping to reduce workload. Teachers and moderators were also impressed with the quality of the six pre-moderated activities they received towards the end of 1998.

Despite the approval of the changes, NZQA lost an opportunity to have a national consistency check as part of the system. Some moderators commented that if the main purpose of the NPA was to check consistency of assessor judgements:

> ... it doesn't have to involve the students i.e. it could be done on some sample work provided by the National Moderator. This could be achieved by supplying a sample of student work to all providers.

The end-point assessor judgement agreement trials reported in section 5.6 perform this function.

5.5 The consistency of moderator decisions related to the moderation of assessment activities

Consistency between moderators is a key quality indicator of the moderation process and was investigated by the research question:

> ***What is the consistency of front-end local moderation of assessment activities?***

There were two approaches to investigating the consistency of moderators' decisions.

The first investigated whether different moderators made comparable decisions when moderating different assessment activities for the same PUS from different schools. This was investigated qualitatively by surveying moderators' opinions on the consistency of moderation.

The second involved running annual moderator agreement trials to investigate whether different moderators moderating the same assessment activity make similar moderating decisions and supply providers with similar feedback.

a) Moderators' opinion on consistency between moderators

In the 1997[31] and 1998[32] questionnaires moderators were asked their opinions on the consistency of moderator decisions between moderators. The results are summarised in Table 5.14.

Table 5.14: Moderators' views on how satisfactory the consistency of moderator decisions was in achieving comparability between schools (1997-1998)

		Percentage of moderators	
		1997	1998
5	Very satisfactory	23	11
4	Satisfactory	50	79
3	Average	23	5
2	Not satisfactory	4	5
1	Very unsatisfactory	0	0
Number of respondents		22	21

Table 5.14 shows that in 1997 and 1998 a large majority (73%-90%) of moderators felt that the consistency was "Satisfactory" or "Very satisfactory". Over the same period a small minority of moderators (23%-5%) felt it was "Average" and only one moderator felt that it was "Not satisfactory". This trend was not tested for statistical significance. [33]

In 1997 one moderator commented:

> I have no idea what decisions other moderators are making. I am working in a vacuum.

In 1998 moderators appeared more confident. One moderator attributed this to:

---

[31] 97MQ3b
[32] 98MQ4d

> The discussions about sample activities at the Local Moderator meetings show that we are all very similar in our approach to moderation and share a relatively common understanding of the standards.

b) The moderator agreement trials (1996-1998)

The second approach to investigating consistency was to run annual moderator agreement trials. These provided an external quantitative indicator of the degree of consistency between moderators and annual results can be compared to investigate longitudinal trends. The trials were held at the Regional meetings. Each year of the study a standard activity was submitted to each of the moderators for moderation. The activities had deliberately inserted flaws. Copies of the assessment activities used in 1996,1997 and 1998 agreement trials and the corresponding moderation pro-formas on which moderator decisions were recorded are contained in Appendices 16, 17 and 18 respectively. Moderators were asked to moderate the assessment activity and record their moderator decisions on the standard NZQA pro-formas.[34] These forms list the moderation criteria against which activities are judged. For each moderation criterion, moderators recorded whether the criterion had been met or not. After the moderation was completed moderators handed in the forms and discussed the activity. A consensus panel approach was used to determine the correct decisions for each criterion. Where consensus could not be reached the National Moderator made a ruling. The pattern of responses for each moderator was compared with the consensus pattern. This enabled percentages of agreement between the consensus decisions and the moderators' individual decisions to be calculated for each moderation criterion. The percentage agreement represents the consistency between moderators and the consensus decision. Moderators make one of three final decisions. The activity may be approved

---

[33] χ^2 analysis not carried out because of insufficient cell sizes.
[34] PHYAM02 IN 1996 and SCIS02 in 1997 and 1998.

for use, be approved subject to modification being made or required to be resubmitted. The consensus decision each year was that the activity should be resubmitted. Each year of the research a global mean and standard deviation of the percentages of agreement was calculated to analyse longitudinal trends in the consistency of the moderation process. Table 5.15 contains a summary of the results.

Table 5.15: Results of the 1996-1998 moderators' agreement trials

	1996	1997	1998
Number of moderators who participated in the trial	14	18	18
Number of moderation criteria	11	7	7
Range of % agreement of moderators on the prescribed moderation criteria	54-85	53-100	58-100
Mean % agreement of moderators	70	80	84
Standard deviation of % agreement	10	16	15
% of moderators who agreed that the activity should be approved	8	0	0
% of moderators who agreed that the activity should be approved subject to modifications being made	23	32	11
% of moderators who agreed that the activity needed to be resubmitted	69	68	89

The percentage of Local and Regional Moderators who participated in the trial each year was consistently high (70% or over) but excluded moderators who did not attend the meetings. Since the National Moderator designed the moderation task, he did not participate in the trial. In 1996, the global average percentage of moderator consistency across all moderation criteria was 70% (standard deviation 10%). This represented a baseline for comparison with percentage consistency of moderation of assessment activities in subsequent years. In 1997, and 1998, this figure increased to 80% and 84% respectively. This trend indicates that moderators became more consistent each year of the study.

A more global indicator of consistency is the percentage of moderators who declined to approve the assessment activity each year. This ranged from 69% in 1996 through 68% in 1997 to 89% in 1998.

The increasing experience of the moderators was also reflected in the time it took to moderate the agreement trial assessment activity each year. In 1996 it took moderators an average of 55 minutes *(SD 27 min)* to moderate a common assessment activity. In 1998 it took moderators on average 46 minutes *(SD 12 min)* to moderate a similar activity with an identical number of deliberately inserted flaws.

c) The approval ratio of assessment activities submitted for moderation

The number of assessment activities that were approved without the need for resubmission each year is another indicator of comparability between schools. This was investigated by the research question:

What is the national approval ratio of assessment activities?

Each year of the study moderators were asked what percentage of assessment activities they approved for immediate use. The mean number of assessment activities submitted to individual moderators that were approved for immediate use increased from 52% in 1996 through 65% in 1997 to 70% by 1998. The standard deviation remained approximately the same (22-25%). This trend was statistically significant ($t_{1996/98} = 3.39$, $df = 27$, $p < 0.01$) and reflects the increasing ability of teachers to design activities that meet the requirements of the PUS.

5.6 Consistency of end-point assessor judgements agreement trials

In 1996 and 1997, agreement trials were conducted to determine the level of consistency of end-point assessor judgements. The aim of these trials was to investigate the research question:

> ***What is the consistency of end-point assessor judgements between different providers?***

The NPAs were selected for carrying out the agreement trials because they were common assessment activities with common assessment schedules. Both years an identical student script for the current NPA was sent to providers. The Teacher in Charge of physics was asked to mark the student script using the assessment schedule supplied. For each performance criterion the percentage of agreement between end-point assessor judgements and the consensus judgements was determined. The consensus was reached after discussing the student work at the regional meetings. The mean percentage of agreement for all of the performance criteria assessed by the NPAs represents the overall level of agreement between providers and the moderators' consensus decisions.

After discussing the agreement trial at the 1996 round of moderator meetings and consulting with colleagues, it was decided that 80% was an acceptable minimum level of agreement between assessor judgements. Therefore the mean percentage of agreement and the number of criteria for which agreement was less than 80% were used as indicators of end-point assessor judgement consistency and to investigate longitudinal trends in consistency.

The 1996 NPA assessed against four elements of PUS 6380[35]. A total of 62 providers requested class sets of the NPA from NZQA. All of these providers were sent a sample student script for marking. Forty-three providers returned the marked student script, giving a response rate of 69%.
The 1997 NPA for Level 2 Physics assessed against two elements of PUS 6382[36]. A total of 99 providers requested class sets of the NPA. A random sample of 50 schools was selected and all of the providers in the sample were sent an identical student answer for marking. The number of providers who returned the marked activity was 29, giving a return rate of 58%.

Copies of the 1996 and 1997 NPAs, assessment schedules, student scripts used for the provider agreement trial and breakdown and analysis of assessor judgements may be found in Appendices 16 and 17 respectively.

Table 5.16 provides a summary of the results of the end-point assessor judgement agreement trials. It shows the number of assessment criteria that were assessed by the NPA each year and the distribution of the percentages of agreement between the moderators' consensus decisions and the provider's assessor judgements across all performance criteria assessed.

[35] *Apply formulae, graphical and vectorial methods to find unknowns for a physical system*
[36] *Demonstrate knowledge of waves*

Table 5.16: Distribution of agreement in the 1996 and 1997 end-point assessor judgement agreement trials

	Number of criteria	
Percentage of assessor agreement	1996 N = 27	1997 N = 23
100	14	12
90-99	7	7
80-89	3	1
70-79	2	1
60-69	0	1
50-59	1	1
< 50	0	0

In 1996, the percentage agreement between assessors ranged from 58% to 100%. The mean percentage of agreement over all criteria was 94% (*SD = 11%*). Full agreement was reached on 14 out of the 27 criteria and over 90% agreement for a further 7 criteria. The level of agreement was less than 80% for only three criteria.

The results of the 1997 agreement trial were similar. The percentage agreement between assessors ranged from 55% to 100% and the mean percentage of agreement over all criteria was 93% (*SD = 12%*). Full agreement was reached on 12 out of 23 criteria and over 90% agreement for a further 7 criteria. The level of agreement was less than 80% for only three out criteria.

The fact that only a few of the criteria registered less than 80% agreement and the mean percentage agreement of end-point assessor judgements was consistently very high indicates that end-point assessor judgement consistency was generally satisfactory. It also illustrates that once the assessment schedule has been submitted for moderation and approved, it can be applied accurately and consistently by assessors.

In 1996 and 1997 the providers were asked to submit six items of student work for verification. NZQA reduced this requirement to four items in 1998 in response to feedback from moderators in their annual reports. The 1998 moderator questionnaire indicated that moderators agreed with this decision. When asked how many items of student work moderators felt was sufficient, the mean number of items of student work recommended was 4.4 (SD 0.8).

5.7 Threats to consistency of moderation

The research question:

What are the threats to achieving consistency through the MAP?

aimed to identify factors which have the potential to adversely affect the consistency of assessment against US and comparability between schools. Some of these factors are policy related, some arise from within the moderation system, whereas still others are provider controlled, such as, the conditions related to assessment, resubmission, reassessment, authenticity and sufficiency. The threats to consistency and comparability identified by moderators in their responses to the annual questionnaires are listed below:

- Assessment and qualification policy uncertainty
- The interpretation of the PUS
- Moderator and provider workload
- Lack of provider training
- The moderator/advisor conflict
- Sporadic provider meeting attendance
- Excessive administrative requirements
- Variability in assessment conditions
- Poor internal moderation
- Insufficient sampling

- Doubts about authenticity of student work
- Variable grounds for resubmission
- Lower reassessment standards

The following section describes each of these threats in greater detail.

Assessment and qualification policy uncertainty

The political uncertainty surrounding the Government Green Paper on future qualification and assessment policy, and the delay of the subsequent White Paper in 1998 led to uncertainty about the future. Some moderators expressed the view that they:

> ... operated in an environment which was characterised by a lack of government support were not even sure of the future of US in their current form or the future of the NZQA.

This uncertainty has led to a "trialling mentality" and a "dabbling with US" which impacted negatively on consistency. Some providers expressed the view that:

> ... since SFC is still the main qualification aimed at by their students, they do not need to be too serious about moderation for the Physics Unit Standards.

The interpretation of the PUS

One of the weaknesses of the moderation process that was mentioned by moderators in 1996 was that:

> ... grey areas in the interpretation of the Physics Unit Standards make it difficult to moderate consistently, when different providers submit assessment activities which interpret the same Unit Standard in slightly different ways.

Requests for PUS interpretation were submitted to the National Moderator who publishes the interpretations in the Physics Moderation Newsletter.

The doubt about the interpretations of the standards was not evident in 1997 and 1998, suggesting that, as moderators become more experienced and high quality assessment exemplars, such as, the NPAs and pre-moderated activities become increasingly available, moderator consistency and confidence improve. The assessment activities in the *Assessment Guide: Physics* were also useful as exemplars in helping both moderators and providers establish consistent interpretations. Some moderators suggested that the assessment activities in the *Assessment Guide: Physics,* should be upgraded to become an item bank of pre-moderated activities.

Moderator and provider workload

The workload associated with moderation was high. Moderators commented that:

> Moderation of assessment against US is very time consuming and moderators are all busy teachers.

The time pressures due to moderation on both providers and moderators created pressure points that at times led to inconsistencies due to rushing the process. These became evident in the check moderation process.

Lack of provider training

Training and socialisation of providers is the first step in the process of establishing consistency. This issue came to the fore in 1997 and 1998 when new providers who had not received training joined the moderation system. One moderator commented:

> Not all assessors have received training and some lack knowledge of assessing against standards and moderation requirements. Some are incapable of designing suitable assessment tasks.

The moderator/advisor conflict

A commonly mentioned strength of the moderation process was that it provided opportunities for professional development. One moderator said that it was:

> ... the best professional development they had had in a long time and it is happening on a massive scale.

The professional support offered by the moderation system was seen as particularly valuable for teachers in geographically isolated areas, teachers who are the only physics teacher in their school and teachers who are new to the profession. Teachers identified the following benefit of this support:

> The professional communication between moderators and providers, the sharing of professional expertise and the continual improvement of assessment.

The moderator contract, however, precluded moderators from engaging in training or professional development of providers as part of their moderator role. One moderator said:

> The system is predominantly postal and more face-to-face contact is needed. My role as a moderator is limited, schools need more training and guidance which I am not supposed to provide.

The above argument indicates that there was a conflict between the roles of moderator and advisor. Some moderators commented:

> ... moderators can and should provide training because the professional development of teachers is one of the most beneficial by-products of the moderation system.

Sporadic provider meeting attendance

Sporadic attendance at local provider meetings poses a threat to comparability. Moderators commented that the:

> Geographical spread of allocated providers and the miserable approach to financing provider meetings, made communication difficult and made it nigh impossible to have local meetings of moderators with providers.

One moderator added:

> ... since these meetings are voluntary and no relief time or travel allowance is made for providers, attendance at the meetings was poor.

In some instances these problems led to no meeting being held. There was widespread support among moderators that:

> There should be NZQA funding for travel for providers to attend meetings with Local Moderators.

Excessive administrative requirements

Moderators complained about the large number of moderation pro-formas (15 in 1998) and suggested that the number of forms needed to be reduced and that the forms need to be redesigned and simplified. The forms were redesigned for use in 1998 but moderators felt that they still required a lot of duplicated unnecessary information to be filled in.

Variability in assessment conditions

The concern about variability in assessment conditions between schools arose because some providers allowed students to work in small groups, others used open book tests whereas others still provided lists of relevant formulae. In the 1998 moderator questionnaire, moderators were asked how satisfied they were that the conditions under which students are assessed were comparable between schools.[37] The responses showed that 65% of moderators were "Satisfied" or "Very satisfied" that the conditions under which assessment occurred were comparable between schools, 25% were "Not sure" and 20% were "Unsatisfied".

One moderator who was "Unsure" commented:

> There appears to be some variation in the amount of time allowed for similar assessments at different schools. Some teachers provide formulae whereas others don't. Some teachers allow open book tests whereas others expect students to remember their notes.

Poor internal moderation

Moderators mentioned that:

---

[20] 98MQ13

> There was no check on the internal moderation procedures between different classes in the same school.

This part of the moderation process is not the responsibility of the moderator, but is meant to be checked internally as outlined in the provider's application for accreditation. Moderators expressed doubts that this was happening adequately in some providers after moderating different assessment activities from different teachers, employed by the same provider.

Insufficient sampling

The sampling of the PUS to be moderated each year was seen as a threat to consistency. Moderators felt that:

> Since only 20% of the programme is moderated each year there is no check on the quality of the other 80%. It may take up to four years before this is finally moderated. This means that it may take up to four years before assessment activities that don't meet the standard are finally picked up

Doubt about authenticity of student work

The procedures used by different physics departments to ensure that the work submitted by students is authentic i.e. their own work, was seen by moderators as a potential source of inconsistency, especially for work completed under uncontrolled conditions, such as projects, extended investigations and practical work in groups. In 1998, 67% of the moderators felt that the moderation process should require evidence of authenticity of student work which is completed under informal conditions and for verification. The reason given for this was:

> Asking students to sign an authenticity declaration prevented cheating and the rampant trading of assignments.

Variable conditions for resubmission

Moderators saw the variability of conditions, under which assessors allowed students to resubmit work for reconsideration, as a threat to consistency. In

1998 moderators were surveyed on how consistent they felt providers were in this respect.[38] The responses showed that 65% of moderators felt that providers were "Average" or "Consistent" in their interpretation of when students are allowed to resubmit work and 35% percent of moderators felt that providers were "Inconsistent".

A moderator who was also a Science Advisor, with a knowledge of assessment procedures at a range of schools, said:

> Resubmission was a potential source of inconsistency both between schools and between teachers in the same school. Some providers do not allow it whereas others go overboard on it.

Lower standards for reassessment activities

The physics moderation system does not require assessment activities that are used to reassess students to be moderated. Consequently there is no check on whether the standard of these activities is comparable to that of the initial assessment. In 1998, moderators were asked whether they felt that reassessment activities should be moderated.[39] Forty-five percent of moderators indicated that it was either "Important" or "Very important" that reassessment activities should be moderated, 30% were "Not sure" and 25% felt that it was "Unimportant" or "Very unimportant".

Moderators gave the following reasons for moderating reassessment activities:

> The moderation process doesn't detect the standard of reassessment. It would be useful to include the reassessment activities in the moderation process, since it is at the reassessment level where the pressure on teachers to 'fudge' the standards is the greatest.

Moderators felt that:

> Reassessment activities should be moderated because they tend to be less formal. Some teachers designed reassessment activities, by

[38] 98MQ14

[39] 98MQ7

> simply changing numbers. This lead to reassessment activities that were easier because students basically got a second go at the same activity.

Insufficiency of evidence

Sufficiency of evidence refers to the number of times a student must demonstrate competency before being eligible for credit for an element. The moderator comments in the 1998 questionnaire showed that there was a range of opinion on what constitutes sufficiency of evidence. Most moderators felt that students had to demonstrate competence either once or twice *(M = 1.5, SD = 0.6)*. One moderator said:

> Mastery surely means that a learning outcome (element) has to be achieved more than once.

5.8 Conclusion

The investigation into the effectiveness of the moderation system produced the following key research findings:

- The annual moderator training programme was successful in helping moderators develop an understanding of their role as moderators, the physics MAP and the moderation of assessment activities and schedules.
- The moderation of the assessment plan, assessment activities and schedules, assessor judgements and check moderation all contributed effectively to achieving comparability between schools.
- Communication in the moderation system was effective.
- The national and regional meetings were satisfactory but the low attendance at the local optional provider meetings was unsatisfactory.
- Teachers and moderators felt that the pre-moderated activities were more useful than the NPAs.
- The consistency of moderator decisions and end-point assessor judgements was consistently high.

Moderators identified a number of barriers to achieving consistency and comparability. These included assessment and qualification policy uncertainty, the moderator and provider workload, the lack of provider training, low attendance at local provider meetings and the moderator/advisor conflict. Furthermore, moderators were concerned that insufficient sampling, poor internal moderation, variability in assessment conditions, variable grounds for resubmission and doubts about the authenticity of student work may adversely affect comparability between schools.

The above research findings and associated recommendations for improving the moderation system are discussed in Chapter 7.

Chapter 6

Manageability of Year 12 Physics Assessment

If a new form of assessment has high validity and reliability but is not manageable for teachers, moderators and students, it will not be practical to implement. This chapter discusses the manageability of assessment against the Level 2 PUS by focusing on the research question:

Is assessment against the Level 2 PUS manageable?

This question was addressed by a longitudinal investigation into teacher, moderator and student workload, the sufficiency and quality of resources, implementation issues, and the impact of assessment against the Level 2 PUS on students and teachers. The manageability of assessment for SFC was used as a benchmark for comparing the manageability of assessment against the Level 2 PUS. The data used to answer this question were derived from annual (1996-98) teacher, moderator and student questionnaires, and five longitudinal case studies.

6.1 Workload associated with Year 12 physics assessment

The manageability of the workload associated with assessment against the PUS is a major factor in deciding whether the new system is practical to implement and maintain. This issue was addressed by the subsidiary research question:

Is the workload associated with administration, moderation and assessment of the Level 2 PUS manageable for teachers, students and moderators?

The workload issues related to these groups are discussed in the following sections.

6.1.1 Teacher workload

To get an objective measure of teacher workload, each year of the study, teachers were asked to list the average amount of time they spent per week on tasks associated with the assessment and administration of the Level 2 PUS.[1] For comparative purposes a parallel question was asked of teachers who assessed for SFC only.[2] The tasks and times are reported in Table 6.1.

Table 6.1: Means and standard deviations of the times spent by teachers each week on tasks associated with assessment and administration of the PUS and SFC (1996-1998)

	Minutes per week											
	PUS						SFC					
	1996		1997		1998		1996		1997		1998	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Writing assessment activities	128	143	79	64	72	13	37	27	34	35	57	100
Marking	104	141	64	79	46	54	78	77	69	59	64	51
Moderation	28	31	17	16	25	35	NA	NA	NA	NA	NA	NA
Department meetings	30	24	24	19	36	52	20	18	10	9	28	33
Maintaining student records	31	19	39	35	38	22	21	16	19	10	74	135
Total	321		223		217		156		132		223	

Table 6.1 indicates that the average amount of time teachers spent each week on tasks associated with assessment against the Level 2 PUS

[1] 96TQT6, 97TQ7, 98TQ7
[2] 96TQT4, 97TQ7, 98TQ7

decreased from 321 minutes (5.4 hours) per week in 1996 through 223 minutes (3.7 hours) in 1997 to 217 minutes (3.6 hours) by 1998. Teachers commented that they found it difficult to estimate weekly averages because of the variable nature of the workload and many left gaps in their responses.

In comparison, the average amount of time teachers spent each week on tasks associated with assessment for SFC varied from 156 minutes (2.6 hours) in 1996 to 132 minutes (2.2 hours) in 1997 and 223 minutes (3.7 hours) in 1998. [3]The low workload in 1997 may be explained by the fact that teachers who practised dual assessment for SFC and the PUS were advised at the cluster meetings to reduce their assessment for SFC to a minimum. Following political uncertainty about the future of the Framework in schools in 1998, many schools had reduced the number of PUS they assessed against and increased their assessment for SFC.

A comparison of the workloads associated with the two systems of assessment was not appropriate in 1996 and 1997 because the initial implementation of assessment against the Level 2 PUS meant that the times reported by teachers were not typical. One teacher suggested:

> It is invalid to compare the workload in implementing the Physics Unit Standards with the workload of assessment for SFC, because resources for the latter were developed over a number of years and now required little maintenance, whereas resource development for the Physics Unit Standards is in full swing.

The most appropriate time to compare the workloads associated with assessment for SFC and the PUS was in 1998. In 1998 most teachers had completed this aspect of the questionnaire. In addition, the workload associated with implementing the new system of assessment could be expected to have settled down and therefore enable a meaningful

[3] Due to the incomplete nature of the data, longitudinal trends in the SFC and PUS workload data were not investigated.

comparison of typical workloads to be made. In 1998 the mean weekly workload associated with assessment against the Level 2 PUS was 217 minutes *(SD = 69)*. The corresponding time for assessment for SFC was 223 minutes *(SD = 80)*. There was no statistically significant difference between these times ($t_{PUS/SFC} = 0.47$, $p > 0.05$).

As expected in the first year of implementation, the biggest contributor to the workload associated with assessment against the PUS, was the writing of new assessment activities and schedules. In 1996 teacher B[4] commented:

> The workload is high due to the writing of new activities, the designing of assessment schedules to match the prescribed format, liaison with the moderator and the administrative demands of recording achievement for every single element for every student.

While marking of PUS assessments was seen as straightforward, the recording of students' achievement was seen by teachers as a time consuming task. For some teachers this involved setting up a revised mark book layout, whereas others set up separate files for each student or developed computerised record keeping systems. The requirement to maintain a portfolio of unit standards assessments for each student also added to the workload.

The moderation requirements also added to teacher workload. Teachers mentioned that they had to prepare new assessments well in advance of the planned assessment date to allow time for pre-assessment moderation. They complained about having to photocopy student work for post assessment verification and the amount of work involved in filling out the various moderation proformas.

[4] See Chapter 3 for an explanation of teacher code references.

The recording and tracking of student achievement was seen as a straightforward but very time-consuming task. All of the case-study teachers kept detailed records of the elements students had achieved for each PUS. Teacher A commented:

> The recording and keeping track of student progress was a huge task and diverted time away from the development of new resources and lesson planning.

To reduce the workload of recording two of the case study teachers had investigated the computerised MUSAC package but had not yet used it. Teacher B had developed his own computerised database and found that:

> Once the computer programme was set up it drastically cut the time required for the recording and reporting aspects of assessing against the PUS. I can get an up-to-date printout of student progress at any time.

All of the case study teachers used a portfolio system for storing student work and keeping track of progress. This involved setting up a filing system with a folder for each student. Special cover sheets were developed to track progress and record dates each element and PUS were achieved. At the end of the year the teachers' records were cross-checked with the portfolios.

In addition to the tasks listed in Table 6.1, teachers mentioned that attending assessment training courses, cluster meetings and the associated setting of relief work further added to the workload.

Case-study teachers' comments reflected that the workload was less in 1997 and 1998. The reason given by teacher B was that:

> NZQA provided an increasing pool of assessment activities such as NPAs and pre-moderated activities and I could re-use activities I wrote last year. I have now built up a good bank of resources.

Teacher D felt that:

> ... reporting and recording became easier now that I am more familiar with these tasks and have established a computer record keeping system for this.

Based on the data presented in this section it may be concluded that three years into the implementation period, the workload associated with assessment against the PUS was similar to that for SFC.

6.1.2 Moderator workload

Since all moderators were engaged in full time employment, it was important that the additional workload of moderation was manageable for moderators. Each year of the study the physics moderators were asked how manageable they found the handling of their moderator workload in addition to their teaching.[5] Table 6.2 summarises the results.

Table 6.2: Moderators' views on how difficult it was to handle the workload of moderation (1996-1998)

		Percentage of moderators		
		1996	1997	1998
5	Very difficult	0	0	0
4	Difficult	54	26	16
3	Average	31	48	53
2	Easy	15	22	21
1	Very easy	0	4	11
Number of respondents		13	23	19

($\chi^2_{1996/1997/1998}$ = 6.89, df = 4, p > 0.05)[6]

[5] 96MQ10, 97MQ9, 98MQ15

[6] Categories 1-2 and 4-5 were combined.

Table 6.2 shows that in 1996, a majority (54%) of the moderators surveyed felt that the workload was "Difficult" to manage. In 1997 and 1998 only a minority (26%-16%) of moderators felt this way. A minority of moderators found the workload "Easy" or "Very easy" to manage (15% -26% -31%). Each year approximately a third to half of the moderators found the workload "Average" (31%-48%-53%). There were no statistically significant longitudinal changes in moderators' perception of the difficulty of handling their workload.

In 1996, one moderator commented:

> Teaching, assessing, pastoral care and extra-curricular activities leave little time for family life let alone moderation.

The workload associated with moderation was compounded by the fact that most of the moderators were involved in implementing the new assessment system in their own schools. One moderator explained:

> There are busy times in the school when you can get overloaded. At one stage I was writing my own exams for Year 12 and 13 physics as well as moderating assessments for other schools. My Saturday evenings are not the same.

Moderators commented that the pressure on their time was not necessarily because of the overall workload of moderation but due to the:

> ... stop/start nature of it i.e. 20 minutes here, 50 minutes there and the requirement to meet tight deadlines for NZQA and providers. Schools often submit work too close to the proposed assessment date and then expect to get it moderated in a hurry.

The moderators who complained of excessive workload generally had large numbers of providers and recommended a maximum of eight providers for moderators who are full time teachers.

While the workload had not significantly decreased in 1997 and 1998, moderators mentioned improved organisation from NZQA, the improvements made to the moderation pro-formas, increased experience and confidence of the moderators and greater provider consistency as reasons for feeling better equipped to handle the workload. One moderator commented:

> The workload was easier to handle this year. I was in a better position to plan my load because all resource material was received well in advance before providers sent me their assessment plans. I also feel more confident about what I am supposed to be doing.

The overall conclusion for this section is that the workload associated with moderation was difficult to manage in 1996. In 1997 and 1998 a majority of moderators found the workload "Average " or easier to manage. This may be because moderators were more experienced.

6.1.3 Student workload

The investigation into student workload used two sources of evidence. The first of these were the opinions of students who had been assessed against the Level 2 PUS and were therefore able to give an informed opinion about how the workload compared with that of their other Year 12 subjects. The second involved a comparison of students' opinion on the number of Level 2 PUS assessments and the number of assessments for other Year 12 subjects.

The data for each of these two student workload indicators were also compared with the workload of students who were assessed for SFC only. This was done to determine whether any differences in workload were due to the nature of the subject or the type of assessment used.

a) Comparison of workload

Every year of the study students were asked how their workload in physics compared to the other SFC subjects they were studying.[7]

[7] 96SQT3, 96SQN3, 97SQ4, 98SQ4 .

Table 6.3 shows a comparison of the workload of students who were assessed against the Level 2 PUS and SFC with the workload of students who were assessed for SFC only.

Table 6.3: Comparison of students' descriptions of the workload in Year 12 physics compared to other Year 12 subjects for students assessed against both the PUS and SFC and students assessed for SFC only (1996-1998)

		Percentage of students					
		PUS and SFC			SFC		
		1996	1997	1998	1996	1997	1998
5	Far greater	2	5	2	2	3	4
4	Greater	16	23	23	20	32	28
3	Similar	60	45	47	59	46	39
2	Less	20	26	26	18	18	28
1	Far less	2	1	2	1	1	1
No of respondents		377	234	163	472	273	337
Mean		3.0	3.1	3.0	3.0	3.1	3.0
Standard deviation		0.7	0.8	0.8	0.7	0.8	0.9

Table 6.3 shows that each year a moderate proportion (60%-45%-47%) of the students who were dually assessed felt that their workload was "Similar" to that of their other Year 12 subjects. A minority of students (18%-28%-25%) felt that it was "Greater" or "Far greater" and approximately a quarter (22%-27%-28%) felt it was "Less" or "Far less". There were no statistically significant longitudinal changes *($F (2, 765)_{1996/1997/1998} = 0.94, p > 0.05$).*

Over the same period, the percentage of students who were assessed for SFC only and responded that their workload was "Similar" was almost parallel (59%-46%-39%). A minority of students (22%-35%-32%) felt that the workload was "Greater" or "Far greater" and a smaller group (19%-19%-29%)

felt it was "Less" or "Far less". There were no statistically significant longitudinal changes $(F\ (2,\ 1073)_{1996/1997/1998} = 2.42,\ p > 0.05)$.

Over the three years of the study an ANOVA revealed a significant difference between students who were dually assessed against the Level 2 PUS and SFC, and students who were assessed for SFC only $(F\ (1,\ 1850)_{DUAL/SFC} = 5.81,\ p < 0.05)$. The students who were dually assessed $(M = 3.0,\ SD = 0.78)$ described the workload in their Year 12 Physics course as slightly less than students who were assessed for SFC only $(M = 3.1,\ SD = 0.80)$. This difference in mean is not practically meaningful but does indicate that the workload for students who were dually assessed was slightly less or similar than for students who were assessed for SFC only.

This was an unanticipated result. Since schools that assessed against the PUS also assessed students concurrently for SFC it might have been expected that students' perception of their workload would have been higher for those students. The above finding shows that this was not the case. In schools where dual assessment was practised this was often done by simultaneously marking the same assessment activities with both a norm-referenced marking schedule and a standards-based assessment schedule. While this is obviously more work for the teacher it does not necessarily mean more work for the students.

Another possible explanation is the difference in the nature of the assessment used. Assessment for SFC typically takes a full class period and traditionally occurs at the end of each topic. Since there are no reassessment opportunities the individual summative assessments carry a large emphasis. Schools that assessed against the PUS typically had a larger number of smaller formative assessments that were carried out close to learning and were often integrated into ordinary lessons. Since students were offered reassessment opportunities the emphasis on individual assessments was less than for SFC. It is possible that students perceived these assessments as part of the learning process and did not regard them as contributing to an increase in workload.

There were no discernible differences in students' comments about workload associated with assessment for the PUS and SFC. The student comments about workload are categorised under the following headings:

- The quantity of work
- Large amount of practical and assignment work
- The difficulty of physics and the rate at which the content is covered
- Homework

Quantity of work

The largest category of responses from both SFC and PUS candidates indicated that the workload of their Year 12 physics course was similar to other subjects. A typical comment made by the minority of students who felt that the workload was too high was:

> There is way too much work for me to possibly get through. There are far more exercises that we are meant to do than in any of the other subjects.

Large amount of practical and assignment work

Students identified the practical nature of the subject as a contributory factor to the workload. One student remarked:

> The workload is higher because of the large amount of practical work that needs to be written up and the assignments, which can be hard.

The difficulty of physics and the rate at which content is covered

Both SFC and PUS candidates commented that physics could be a difficult subject to understand. Closely allied to this were comments about the fast rate at which the content of the subject is covered in class. One student said:

> Physics is hard to understand. My teacher moves quite quickly and you have to keep up with the work. If you don't understand first off, you don't get the chance to learn. You need to be good at mathematics.

Amount of homework

Although homework was mentioned by a large number of students, there was little agreement on how much homework there was compared with other subjects. There seemed to be a general trend for teachers to set aside class time for students to complete assignment work in class because many students said they could not do the problems unaided and needed teacher assistance. As one student said:

> Most of the work is completed during class time so there is not a lot left for homework. Sometimes when you work alone you get stuck. I prefer doing the work in class where the teacher can help us.

b) Number of assessments in Year 12 physics

Another indicator of student workload is the number of assessments in Year 12 physics. Each year of the study, students were asked how the number of assessments (exams, tests, assignments, etc.) for physics compared with their other Year 12 subjects.[8]

Table 6.4 compares students' opinions on how the number of assessments in Year 12 physics compared with the number of assessments in other Year 12 subjects. This is broken down into the opinions of students who were dually assessed against the Level 2 PUS and SFC students who were assessed for SFC only.

---

[8] 96SQT4, 96SQN4, 97SQ5, 98SQ5

Table 6.4: Comparison of students' descriptions of the number of assessments for physics in comparison with other Year 12 subjects for students assessed against the PUS and SFC and students assessed for SFC only (1996-1998)

		Percentage of students					
		PUS and SFC			SFC		
		1996	1997	1998	1996	1997	1998
5	Far greater	8	9	5	4	4	5
4	Greater	41	37	28	26	41	25
3	Similar	37	39	50	56	37	46
2	Less	12	15	16	13	17	23
1	Far less	2	0	1	1	1	1
No of respondents		377	234	163	472	273	337
Mean		3.4	3.4	3.1	3.3	3.3	3.0
Standard deviation		0.8	0.9	0.8	0.7	0.8	0.8

Table 6.4 shows that in 1996 and 1997, approximately half (49%-46%) of the students who were dually assessed against the Level 2 PUS and SFC, found that the number of assessments for physics was "Greater" or "Far greater" than their other Year 12 subjects. In 1998 only a third (32%) of students felt that this was the case. Over the same period, there was an increase in the proportion of students who felt that the number of assessments was "Similar" to their other subjects from approximately one-third (37%-39%) in 1996 and 1997 to half (50%) by 1998. The percentage that felt that it was "Less" or "Far less" remained steady at about 15%. These trends were statistically significant $(F\ (2,\ 765)_{1996/1997/1998} = 5.23,\ p < 0.05)$. The post hoc Scheffé analysis showed that the number of assessments was similar in 1996 and 1997 and decreased in 1998 $(p < 0.001)$.

Students who were assessed for SFC only responded differently. Each year a smaller percentage (30%-45%-30%) of students felt that the number of assessments for physics was "Greater" or "Far greater" than their other subjects and a larger group (56%-37%-46%) felt it was "Similar". There were no statistically significant longitudinal effects *($F\ (2,\ 1073)_{1996/1997/1998} = 2.42,\ p > 0.05$).*

An ANOVA revealed a very significant difference between students who were dually assessed against both the Level 2 PUS and SFC and students who were assessed for SFC only *($F\ (1,\ 1847)_{PUS/SFC} = 10.30,\ p < 0.005$).* The former *($M = 3.4,\ SD = 0.8$)* reported a slightly larger number of assessments than the students who were assessed for SFC only *($M = 3.2,\ SD = 0.8$).*

Students who were dually assessed commented on the increased frequency but smaller size of PUS assessments. This is illustrated by the following comment:

> The assessments we get are quite easy; we have shorter assignments like lab reports which is good, but we get them every three weeks. I don't mind because the more frequent tests mean they are not so long and less work to concentrate on at once.

Another student said:

> We have a larger number of smaller tests in physics that have less weighting than in other subjects but they are much more closely related to what we are doing in class at the time.

Some students complained about the number of assessments and felt that there was too much emphasis on assessment. One student explained:

> I think the course is too focused on assessment and less time is available for focusing on learning. We always seem to be preparing for another test.

The main findings related to student workload were:

The workload for students who had been dually assessed against both the Level 2 PUS and SFC was perceived to be similar to that of their other Year 12 Subjects and slightly less than that of students who were assessed for SFC physics only.

Students who were dually assessed against the Level 2 PUS and SFC reported a comparatively slightly greater number of assessments compared with their other SFC subjects than students who were assessed for SFC physics only. There was however a statistically significant reduction in the number of assessments over the period of the study.

The overall conclusion is that the workload for students who were dually assessed against the PUS and SFC was similar to that of students who were assessed for SFC only.

6.2 Resources and implementation support

Sufficient quality resources must adequately support implementation of a new system of assessment. The issue of resourcing was focused on by the research question:

> ***Are the resources and support provided by NZQA sufficient for the implementation of the Level 2 PUS?***

The NZQA provided the following resources and support to assist teachers in the implementation of the PUS:

- Quality and format of the PUS
- The *Assessment Guide: Physics.*
- Teacher training in standards-based assessment
- The cluster meetings.

6.2.1 The quality and format of the PUS

The investigation addressed the following aspects of the quality and format of the PUS:

- The quality of the first edition of the PUS used in the 1996 trial
- The quality of the revision of the PUS that occurred at the end of 1996
- The preferred format of publication and accessibility of the PUS

a) The quality of the first edition of the PUS

In the 1996 questionnaire, teachers were asked to comment on the format and quality of the first edition PUS. They identified the following strengths:

- The Level 2 Physics Unit Standards closely matched Level 7 of *Physics in the New Zealand Curriculum.*
- The elements of the Level 2 Physics Unit Standards were appropriate and important learning outcomes for Year 12 physics.
- The performance criteria were specific and enabled accurate marking of student work.

Teachers mentioned the following aspects that needed to be improved:

- Large Unit Standards need to be split up.
- The range statements are too cumbersome and need to be simplified and reduced.
- There are too many performance criteria. These need to be reduced and simplified.
- The credit values need to be adjusted to reflect actual classroom teaching time.
- The wording of the Unit Standards is ambiguous in places.
- Criteria for excellence should be built into the Physics Unit Standards.

b) The quality of the 1996 revision of the PUS

In term four of 1996, the author was contracted to revise the Level 1-3 PUS. The main aim of the revision process was to address the areas of improvement identified by teacher feedback and to make assessment against the PUS more manageable. Apart from the inclusion of criteria for excellence, all of the recommendations made by teachers in Part a) were implemented.

In the 1997 questionnaire, teachers were asked whether they thought the revised PUS were an improvement on the PUS used in the trial.[9] The responses showed that 64% of teachers felt that they were a "Definite improvement", 33% were "Not sure" and only 3% felt that they were no improvement. No teacher felt they were "Worse".

Teachers expressed support for the review:

> It is good to see the review process working. The wording of the elements has been simplified and they are easier to interpret, less ambiguous. I find them easier to manage and understand, and not as picky as the earlier ones. It was sensible to reduce the number of performance criteria and to split the electricity and magnetism into separate units.

Teacher A felt that the revised PUS were considerably simplified but felt concerned about the withdrawal of US 6384. He felt that the generic replacement US was more suitable for biology and chemistry than physics.

The splitting up of larger PUS into smaller ones was viewed with uncertainty. One teacher commented:

> Some of the Physics Unit Standards contained a whole lot of elements and performance criteria that made them clumsy and unwieldy and these have been simplified. This made them easier to administer.

[9] 97TQ12

> However simplification meant that you end up with smaller chunks of work which made the standards even more fragmented. There is clearly a conflict between these aspects.

c) The published format of the PUS

The responses to the 1996 teacher questionnaire indicated that teachers felt frustrated that they did not receive a personal copy of all the relevant PUS. In 1997 they were asked what format they preferred for receiving a copy of the PUS from NZQA.[10] Fifty-four percent of teachers wanted a personal paper copy, 32% preferred downloading specific PUS from the *Framework Explorer* CD when needed and 14% wanted both a paper copy for use at home and the CD for use at school.

The following comment was typical of teachers who preferred a paper copy:

> I need a paper copy of all the Physics Unit Standards that are relevant to my Year 12 and 13 Programme. NZQA should provide this, just like the Ministry of Education provided a personal copy of the physics curriculum. I need it in a form that is accessible at home and during holidays to enable me to plan and write assessment activities. I don't have time during the school day for the lengthy process of printing stuff off. A paper copy can be read at home at leisure and can easily be photocopied for other members of staff.

Teachers who preferred the *Framework Explorer* CD commented on its high quality and the convenience of having an electronic copy that they could subsequently adapt for their own use. Others identified limitations to using the CD. One teacher said:

> The *Framework Explorer* CD is difficult to use and not user friendly for people who do not have computer skills. It should be more user

[10] 97TQ24

> friendly and editable. I want to be able to cut and paste from it to insert titles and elements directly into my own assessment activities.

Problems with the expiry date meant that teachers were not able to access the CD after the expiry date, even though the new CD had not yet been delivered to schools.

There was also a problem with access to facilities to use the CD. One respondent said:

> We don't all have computers with CD drives at our fingertips. We do have CD facilities at my school, but it is not accessible in my work area. Why not supply a floppy of the Physics Unit standards?

This was followed up and floppy discs of assessment activities and US became available later in the year.

The overall conclusion for this section is that while teachers had criticisms of the first edition of the PUS, 64% felt that the second edition was an improvement. Fifty-four percent of teachers wanted a personal paper copy of the PUS and while the CD was thought to be of high quality there were problems associated with accessing it.

6.2.2 The *Assessment Guide: Physics*

The *Assessment Guide: Physics* was designed to assist teachers with all aspects of assessing against the PUS. It contains a sample school scheme which shows how the PUS can be incorporated into a school programme, a section on how to write assessment activities and schedules, an item bank of assessment activities for Levels 1-3, and information on the moderation process, reassessment and resubmission, portfolio keeping, recording of results and reporting. The first version of the guide was available to physics teachers who participated in the 1996 Level 2 PUS trial. Each year teachers who had assessed against the PUS were asked how useful they found the *Assessment Guide: Physics* as an aid in establishing the difficulty level of assessment activities they wrote themselves.[11] Table 6.5 contains a summary of the results.

Table 6.5: Teachers' opinions on the usefulness of assessment activities in the *Assessment Guide: Physics* (1996-1998)

		Percentage of teachers		
		1996	1997	1998
5	Very useful	42	13	16
4	Useful	38	28	60
3	Not sure	0	26	9
2	Of limited use	19	18	13
1	No use	0	15	2
No of respondents		26	39	55

($\chi^2_{1996/1997/1998}$ = 31.79, N = 120, df = 8, p < 0.001)

Table 6.5 indicates that the percentage of teachers who felt that the activities in the *Assessment Guide: Physics* were either "Useful" or "Very useful" decreased from 82% in 1996 to 41% in 1997 and subsequently increased to 76% in 1998. The lower rating in 1997 was statistically significant and may be explained by the fact that the 1996 bank of activities became obsolete following the 1996 revision of the PUS. In 1998 a CD of updated activities was distributed to schools which were generally of high quality. This explains the increase in teachers' perceptions of the usefulness of the activities in 1998.

The following comment made in 1996 shows why teachers felt that the assessment guide was useful:

> The assessment guide was a good source for ideas and saved hours and hours of workload. It was especially useful as an item bank for reassessment activities. It helped me to interpret the standards and

[11] 96TQT7d, 97TQ18, 98TQ18

> I relied extensively on it during the year.

Teacher D saw the *Assessment Guide: Physics* as a good source of assessment activities:

> I used much of the material in the guide. It was an essential resource in a sole teacher department like mine. To remain useful the guide needs to be kept up to date and continually improved. The section on record keeping in the *Assessment Guide: Physics* provided some good ideas on how to set up a recording system, and the assessment activities provide a useful guide for teachers to write their own activities.

Teacher A commented on the usefulness of the implementation suggestions in the guide:

> The record keeping pages gave me some useful ideas but the assessment activities included in the guide were based on the draft Physics Unit Standards. There was little material included for the US we attempted early in the year.

In 1997 the proportion of teachers who felt that the activities were "Very useful" or "Useful" reduced to 41% *($p < 0.001$).* Several teachers expressed the need for more activities to match the revised PUS and wanted the activities in the guide to be pre-moderated and the mistakes to be edited out. One teacher commented:

> It would be good if the moderation system was used to have each activity looked at by a number of moderators. This would improve the quality of the guide and be useful in establishing consistency of moderation.

In 1998 teachers were provided with an updated version of the *Assessment Guide: Physics* which contained a batch of new activities which were written at a five-day writing party. Teachers responded that the assessment activities were of a higher standard than the assessment activities in the original assessment guide and included activities that had been trialled in schools in 1996. One teacher commented:

> The activities in the first batch were not moderated and untried. The later ones we got through the trial were much better. I found some gaps but could adapt other activities to the Physics Unit Standard I was assessing.

6.2.3 Teacher training in standards-based assessment

Prior to participation in the 1996 Level 2 PUS trial, physics teachers were offered three days training in standards-based assessment. In 1996 they were asked how useful they found the 3-day training programme.[12] The responses show that 96% of teachers felt that the three-day training programme was either "Useful" or "Very useful" and only 4% felt it was "Of limited use".

The following comment illustrates the reason for its usefulness:

> The three-day training programme was a unique and long overdue opportunity for professional development. Without the training I wouldn't have any idea of what was expected. The training was the start of the implementation process and let us know what we were getting involved in. It would have been impossible to get going without it.

Another teacher remarked:

> It helped me clarify my own ideas on how my department would integrate the assessment into our overall programme. The training cleared lots of misunderstandings on day one and the other days provided useful practice in the writing of assessment activities and marking of student work.

[12] 96TQT7f

Several teachers commented that forward planning and preparation was difficult because the training used draft PUS that were different from the PUS that were registered the following year.

All of the case-study teachers received three days training and although opinions about some aspects varied they found it to be a useful introduction to standards-based assessment.

A minority of teachers from some regions commented that they had missed out on training because there were not enough schools assessing against the PUS in their region. New providers who commenced in 1997 and 1998 had insufficient training opportunities available to them and were referred to Advisory Services.

Overall it may be concluded that the three-day training programme was highly successful and should have been continued in the same format in the following years.

6.2.4 The cluster meetings

Cluster meetings were held throughout 1996 trial year, to provide teacher support for the implementation of the Level 2 PUS trial. In 1996, teachers were asked, how useful they found the cluster meetings.[13] The responses indicated that 96% of trial-school teachers found the cluster meetings either "Useful" or "Very useful" and only 4% felt that they were "Of limited use". The most useful aspect of the meetings referred to, was the contact with other physics teachers and the sharing of ideas. One respondent stated:

> The best way to solve problems is still discussion and personal contact with others in the field especially in small schools with only one physics teacher in the department.

[13] 96TQT7e

Teachers generally saw the meetings, as a valuable forum where solutions to common difficulties were discussed and shared. One teacher said:

> ... the cluster meetings were professionally run and I could not have survived the trial without the professional guidance and support provided by the NZQA staff.

Teacher A commented:

> The meetings focused mainly on resolving queries and answering criticisms about US. It would have been more productive to share resources.

Teacher E supported this view:

> Too much time was wasted by people who used them as a moan session to push their own anti US views and that more time should have been spent on positive aspects, such as the writing and sharing of resources.

Teacher B felt:

> ... that the cluster meetings that were held only in 1996, should have been continued in subsequent years.

Overall it may be concluded that the cluster meetings were highly successful and should have been continued in subsequent years.

6.3 School-based implementation issues

The annual interviews with the TIC of Physics at five Canterbury schools were used to address the following research question:

> ***What are the school and physics department based issues related to the implementation of assessment against the Level 2 PUS?***

Most of the issues raised by the case-study schools did not differ from those identified by the teacher questionnaires. These comments are not repeated here because the case studies did not provide any additional information to that already presented. The following issues are reported because the case-study data provided more detail than that obtained by the questionnaires:

- dual assessment
- resubmission and reassessment
- sufficiency of evidence
- authenticity of student work
- communication
- moderation

6.3.1 Dual assessment

Throughout the period of the study the majority of schools which assessed against the Level 2 PUS continued to assess for SFC. This practice of dual assessment created extra work related to writing assessment activities with dual marking schedules, submitting activities for moderation, marking, recording, reporting and administration involved in running two parallel systems of assessment. Four of the case-study schools assessed for the PUS and SFC simultaneously by using the same assessment activities but applying two different marking schemes. School B used separate assessment activities for SFC and the PUS. Schools developed different strategies to minimise the workload of dual assessment. Teacher A explained how this operated:

> We constructed tests with tick boxes for the performance criteria on the side. It was a matter of ticking the boxes for US credit and adding up the ticks at the end to generate a SFC grade. This way we had to mark once only.

Dual assessment was seen to be confusing for students. Teacher D commented:

> The dual assessment for both the Unit Standards and SFC was awkward at first and took up a lot of time but this improved once a system was set up. Students were unclear about what qualification to go for. Since only a few other subjects in the school assessed against

> Unit Standards, it made it practically impossible for students to achieve the required number of credits for the National Certificate. Consequently not all students had hooked onto the Framework. The majority of students regarded SFC as the most relevant qualification.

6.3.2 Resubmission and reassessment

Resubmission is the process where the teacher allows the students the opportunity to correct minor mistakes and to resubmit their script with the mistake(s) corrected. Teacher A explained:

> Resubmission is allowed in cases of minor mistakes or omissions such as missing a unit or a calculation that is done incorrectly. I do not draw students' attention to the mistake and mark the student scripts in such a way that students cannot directly identify the mistakes.

The assessment policy of the school of Teacher B defined resubmission as new evidence provided by the student for the same assessment task so that the teacher can make a final pass or fail judgement. Teacher B explained how this worked in practice:

> Where students have not met the standard because of minor mistakes or omissions they are allowed to resubmit their work together with the corrections so that they can be awarded credit. When the tests are returned to the students they can look at it, and then resubmit it straight away before anyone talks about it or anything. So each test we let them do that when we first gave it back. The sorts of things that may be resubmitted, are left out units or something, but they have to find out for themselves. We don't highlight the areas or anything.

Teachers at the case-study schools were in agreement about the types of incomplete student evidence that constituted grounds for resubmission. All teachers mentioned missing or incorrect units, omissions and minor errors. In practice however they found it difficult to apply this interpretation consistently. Teacher A raised the question:

> How do you compare one major error with a larger number of minor errors?

Opportunities for reassessment were provided at a later time and students were reassessed using a different assessment activity. Some schools set aside regular times during study periods whereas others had a period at the end of the year set aside for reassessment. Still others used the end of year exam as a reassessment opportunity. None of the sample schools carried out reassessment out of normal time-tabled hours. All case-study teachers felt that reassessment was one of the main factors that contributed to excessive workload and wanted to limit the number of reassessment opportunities available to students to one or two.

Prior to commencing the 1996 trial school A had a clear policy in place on reassessment. Teacher A commented:

> We had a policy that we would only allow one reassessment per student per assessment activity. We did not want it to be an indefinitely extendible process where students would keep coming back to us. We could not manage that. So we decided to limit it to one. We had a programme in place where assessment could fit in with the school's assessment process. We thought for example that the time for reassessment could be at the end of the year during the three-hour exam. It was a sort of flow over process that we were looking at.

To reduce the number of reassessment events at school B, assessment activities were constructed that give students multiple opportunities to present evidence that they had met the performance criteria. This reduced the number of students that needed to be reassessed.

Finding time to do reassessments, especially when students were absent posed challenges with some schools resorting to form meetings or lunchtime. This approach cannot be used when more subjects are assessed against US and the demands on available time increase.

In 1998 students were asked how useful they found the opportunity for reassessment. The responses indicated that 74% of students felt that reassessment was either "Useful" or "Very useful", 11% were "Not sure", and 15% felt it was "Of limited use".

6.3.3 Sufficiency of evidence

Sufficiency of evidence refers to the number of times that students have to demonstrate that they have met all of the performance criteria for an element before they can receive credit. Four of the case-study schools required students to demonstrate that they had mastered the performance criteria for credit of an element once, whereas school C required students to demonstrate this twice in the same assessment activity. Teacher E commented:

> One demonstration is sufficient because prior to a formal assessment I have usually seen the students meet the criteria several times in class. In fact I do not assess the students till I am confident that they are ready. This also reduces the number of instances that students need to be reassessed.

Teacher B used assessment activities that addressed sufficiency by providing multiple opportunities for students to meet some performance criteria and specifying that the students had to meet the requirements two out of three times.

6.3.4 Authenticity of student work

Authenticity is the term used to describe that the work which is submitted by the student for marking is the student's own work and is completed without outside or unacknowledged assistance. Authenticity is easy to check in a formal test situation in a classroom setting. It is more difficult to police where student work is carried out under informal conditions such as a take-home project or investigation, or in the case of practical work that is carried out in small groups. All of the sample schools addressed this issue by requiring students to sign authenticity declarations as a check that such work was completed without unacknowledged outside assistance. The assessment policy of school C states that students must ensure that work presented for assessment is their own work and not copied from other students or the result of unacknowledged assistance by persons beyond the school.
Teacher C explained:

> The kids have to guarantee that it is their own work. If they cannot it is cheating and we would confront the kids.

Teacher B said that more was involved than the students signing a piece of paper.

> A lot of it involves going around the students and being aware of what they are doing and what progress they are making. If someone comes along the next day and they have made dramatic progress, cheating is a real possibility. If I thought someone had cheated I would investigate further by asking them detailed questions about their work.

6.3.5 Communication

Each year of the study physics teachers who assessed against Unit standards were asked "How useful was the communication you received from NZQA in explaining the administrative procedures associated with assessment and moderation of the PUS?"[14] The responses are summarised in Table 6.6.

Table 6.6: Teachers' perceptions of the usefulness of communication from NZQA (1996-1998)

		Percentage of teachers		
		1996	1997	1998
5	Very useful	32	10	10
4	Useful	45	49	49
3	Not sure	9	19	11
2	Of limited use	14	14	24
1	No use	0	8	6
No of respondents		26	49	49

($\chi^2_{1996/1997/1998} = 8.52$, $df = 6$, $p > 0.05$)[15]

Table 6.6 shows that each year a clear majority (77%-59%-59%) of teachers felt that communication from NZQA was either "Useful" or "Very useful". Over the same period, a minority (14%-22%-30%) thought it was "Of limited " or "No use". There were no statistically significant longitudinal effects.

Particularly positive reference was made to the help line set up for trial schools and the trial newsletters which "contained up to date information". Teachers commented that the information received from NZQA was comprehensive and they described positively their relationship with their local moderator. Some complained of the information overload.

In 1997 teachers mentioned additional issues related to communication. The internal school communication systems were seen to be responsible for time delays or information getting lost. One typical comment was:

[14] 96TQT8g, 96TQN6f, 97TQ23, 98TQ23
[15] Categories 1 and 2 were combined.

> It is hard to keep up to date with everything when some information gets lost by the HOD and other information is part of more general stuff the NZQA liaison teacher sits on.

Another category of comments related to the verbose nature of NZQA circulars. An NZQA liaison teacher commented:

> Most NZQA communications are gobbledegook. I am an NZQA Liaison officer so I have first hand knowledge of their wordy tomes. There is too much paper. I don't have the time to plough through reams of garbage seeking the nuggets of useful information. My knowledge has been gained by osmosis from other staff that had to suffer to gain what they know.

Another teacher said:

> I need more backup from NZQA. Clearer guidelines from NZQA on timelines etc. and direction would be helpful i.e. will US stay or be dumped? Being at an isolated school also makes it difficult especially being the only physics teacher in the school.

The communication of assessment results to NZQA was seen as cumbersome because:

> You have to submit a detailed record of all the Physics Unit Standards that have been achieved by each student. For SFC all you have to submit is a single grade per student.

6.3.6 Moderation

The case-study teachers felt that the requirement to submit all assessment dates at the commencement of the school year was unrealistic. Teacher A commented:

> Dates may vary because of classes of different ability covering the work at different rates. If you want to change the dates you have to liaise with the moderator. That is a real pain!

The moderation of assessment activities was seen as crucial to the credibility of the system and valuable as professional development. The case-study teachers valued the feedback and liaison with moderators and found the verification of assessor judgements less valuable. The photocopying of student work was seen as a time consuming task. All of the case-study teachers agreed with the 1998 reduction of the number of items of student work which have to be submitted for verification from six to four.

The moderation pro-formas that have to be completed when submitting work for moderation were seen as cumbersome and required needless repetition of information. The forms were not thought to be very user-friendly.

All of the case-study teachers said that they had had a good relationship with their Local Moderator and were satisfied with the turnover period of material submitted. The moderators generally returned material submitted for moderation within 1-2 weeks. There was a perception that different moderators may have had different standards and that some moderators were more likely to approve activities than others.

In 1998 a new position of co-ordinating moderator was added to the moderation system to co-ordinate the assessment plans between the various sciences. The teachers felt the role of co-ordinating moderator created confusion and welcomed the deletion of this role in 1999.

6.4 Conclusion

This chapter presented the following main research findings related to the manageability of assessment against the Level 2 PUS.

Assessment against the Level 2 PUS was manageable for teachers, moderators and students. Students who were assessed against the Level 2 PUS reported a greater number of assessments than students who were assessed for SFC only but the effect of dual assessment may have had an impact on this finding.

Overall the resources and support provided by NZQA was generally satisfactory. While teachers had criticisms of the first edition of the PUS, 64% felt that the revised PUS were an improvement. The majority of teachers wanted a personal copy of the PUS and some encountered problems accessing the Framework Explorer CD. Ninety-six percent of teachers valued the cluster meetings and thought that the *Assessment Guide: Physics* was a useful resource.

School based implementation issues raised by teachers related to dual assessment, resubmission, reassessment, sufficiency of evidence, authenticity, communication, the NPA and administration.

Chapter 7 discusses the research findings related to the manageability of assessment against the Level 2 PUS and makes recommendations for making the system more manageable for teachers, students and moderators.

Chapter 7

Discussion and Recommendations

The paradigm shift from norm-referenced external examinations to internal competency-based assessment is neither complete nor fully endorsed by teachers (§1.4). This was illustrated by the considerable debate that accompanied the introduction of assessment against US in the senior secondary school and culminated in industrial action by the NZPPTA in 1996. The debate and subsequent inquiries and reports highlighted unresolved issues that generally related to the quality management of assessment (§1.3). The literature review established validity, reliability and manageability as traditional key criteria for investigating the quality of assessment (§2.1). These criteria were reconceptualised for evaluating the quality of assessment against the Level 2 PUS and provided a structure for the formulation of the research questions (§2.8). This chapter discusses the research findings, places them in the context of the research literature, formulates overall conclusions and makes research-based recommendations in relation to these conclusions. To illustrate the links that exist between the research findings and the recommendations, the latter have been interspersed with the discussion rather than presented in a separate section. A final section discusses the limitations of the present research and suggests possible avenues for future study.

7.1 Validity of assessment against the PUS

After a review of the NQF, Smithers (1997) concluded that assessment against US should be abandoned because its validity, consistency and fairness could not be ensured. The present research addressed this global concern about validity with a detailed investigation into the curriculum fidelity (including content and construct validity), concurrent validity, validity of the reporting process and consequential validity of assessment against the Level 2 PUS. The following sections discuss the research findings and associated recommendations related to each of these aspects of validity.

7.1.1 Curriculum fidelity

For assessment against the Level 2 PUS to have satisfactory curriculum fidelity, the PUS must address all of the content and skill objectives of the physics curriculum at an appropriate level. In addition the credit allocations must be appropriate, representative of curriculum emphases and proportional to time allocations in school programmes.

The research found that each year a substantial majority of teachers who had assessed against the Level 2 PUS felt that the PUS represented Level 7 of *Physics in the New Zealand Curriculum* well. Teachers commented that the PUS enabled them to assess the full range of Level 7 curriculum content appropriate for a Year 12 physics course and that they mapped well onto existing school schemes (§4.1.1). These findings indicate that Smithers' (1997) general concern about the difficulty of assigning meaningful levels to US was largely unfounded in the case of the Level 2 PUS. There were some minor exceptions to this. A minority of teachers was concerned about the presence of some Level 6 curriculum content in the Level 2 PUS and others expressed disagreement with the deletion of PUS 6384 because the generic science US that replaced it was considered to be too general for a physics course. These are realistic validity concerns that can be addressed by the following recommendations.

Recommendation 7.1
The content of the Level 2 PUS should be restricted to Level 7 of "Physics in the New Zealand Curriculum".

Recommendation 7.2

The investigation PUS 6384 should be reinstated.

The implementation of these recommendations would enhance the content validity of assessment against the Level 2 PUS by strengthening the links with Level 7 of *Physics in the New Zealand Curriculum*.

A number of commentators have expressed concern that Unit Standards for academic school subjects may not be able to be expressed unambiguously (Irwin, 1995; Irwin, Elley and Hall, 1995; Milne, 1996 and Smithers, 1997). This concern has also been raised specifically for the PUS (Batchelor, 1996; NZIP, 1996). This concern was not specifically addressed by the present research. However, over the period of the research there was only a small number of requests for PUS interpretations submitted to the National Moderator, indicating that the elements and associated performance criteria were generally well defined, unambiguous and specific. A related finding was that each year a substantial majority of teachers found the performance criteria "Useful" or "Very useful" for describing students' achievement (§4.3.3.1). Teachers felt that the performance criteria were useful because they "showed up exactly what skills students have mastered". They commented that the specificity of the criteria contributed to a high degree of reliability in marking. These findings indicate that the Level 2 PUS meet Coogans' (1996) criterion that US must be unambiguous for learners and assessors.

A negative consequence of the precisely defined elements and performance criteria was that teachers perceived that the PUS were narrow, prescriptive, specific and rigid and did not allow for creativity in question design or unusual and innovative student approaches to problem solving. They felt that the criteria did not enable recognition of student answers that are correct but use unanticipated strategies or are presented in an unusual format. This echoes concerns that assessment against US can stamp out creative flair by being defined solely in terms of discrete, observable and measurable behaviours (Marshall, 1994; Irwin, 1995, Salter and Hayden, 1996 and Sinclair, 1997). There is clearly a tension between the requirements for the PUS to be unambiguous, specific and clearly defined while at the same time broad enough to allow creativity, innovation and flexibility in question design.

A possible solution to this dilemma is to revise the PUS by writing broader performance criteria. While this suggestion has the potential to adversely affect validity, any resultant loss of specificity can be addressed by expanding the "special notes"[1] in the PUS to clearly prescribe and delineate the content that may be assessed.

Recommendation 7.3

Future revisions of the PUS should broaden the performance criteria to allow for more flexible assessment approaches and variety in the presentation of students' answers. This should be accompanied by the addition of special notes that clearly prescribe the content that may be assessed.

This recommendation was partly implemented in the 1996 revision of the PUS and is illustrated by the revision of PUS 6380[2]. Element 2 consisted of five performance criteria. In the revision this was reduced to three broader performance criteria. These new performance criteria related to the same content and skills but allowed greater flexibility in the design of assessment activities and professional judgement in the marking of students' answers. Sixty-four percent of teachers who assessed against the PUS in 1997 felt that the revised PUS were an improvement over the PUS used during the trial (§6.2.1) indicating that this type of revision was supported by teachers and could be implemented more widely. While it appears unlikely that the tension

[1] The special notes provide additional details of the content to be assessed and how this relates to the curriculum.
[2] *Apply formulae, graphical and vectorial methods to find unknowns for a physical system.*

between the requirement to be unambiguous while allowing for creativity and flexibility can be fully resolved this approach may be useful in future revisions. Concerns that assessment rather than the curriculum might end up driving teaching programmes (NZPPTA, 1997a) and consequently adversely affect curriculum coverage (Smithers, 1997), were allayed by the finding that schools that dually assessed against the Level 2 PUS and SFC achieved a high degree (approximately 90%) of curriculum content coverage that was similar to schools that assessed for SFC only (§4.1.1). This indicates that schools could meet the additional demands of assessment against the PUS without adversely affecting curriculum coverage. Furthermore an investigation into the validity of the Level 2-credit allocation showed that it was generally accurate and in proportion to the amounts of class time teachers traditionally allocated to teaching the related topics (§4.1.2). There were some minor exceptions to this. The statistical analysis and specific comments made by seven respondents indicated that they considered the credit value of PUS 6378[3] to be insufficient. In addition the analysis indicated that the credit value of PUS 6384 was not representative of the class time teachers allocated to the extended investigation. However teachers did not refer to this in their comments. This may be because PUS 6384 assessed an aspect of the Year 12 physics course that was first introduced in 1996 by the new curriculum and had not yet been fully implemented by some assessors. Therefore the comments related to PUS 6384 may indicate a lag in practice rather than an inaccurate credit allocation. The implementation of the following recommendation would enhance the representativeness of the credit distribution.

Recommendation 7.4
The credit value of PUS 6378 should be increased to 6 credits.

Since the credit distribution of the PUS was generally accurate and teachers were able to specifically suggest minor modifications, Smithers' (1997)

[3] *Demonstrate knowledge of motion in one and two dimensions.*

concern that credit values for US may not be able to be assigned meaningfully is unfounded in the case of the Level 2 PUS.
In addition to content validity, the research also addressed aspects of the construct validity of the Level 2 PUS. Commentators' expressed doubt that all learning outcomes can be adequately assessed using a US approach (Batchelor, 1996: NZIP, 1996 and Salter and Hayden, 1996). This was addressed by investigating whether the PUS were suitable for assessing a number of key skills related to the achievement objectives of the curriculum. The skills investigated were practical work, problem solving and the explanation of physics concepts.

The research found that at least two-thirds of teachers who had assessed against the Level 2 PUS found them appropriate for assessing practical work skills (§4.1.3a). This finding concurs with a finding of the NZIP (1996) survey that 66% of physics teachers who had assessed against the Level 2 PUS felt that assessment against the PUS was appropriate for assessing practical skills. Teachers cautioned however that assessors needed to be careful about ensuring the authenticity of student work when practical work was carried out in small groups.

The NZIP survey also found that 60% of the teachers who had assessed against the PUS felt that they were inappropriate for assessing higher order thinking skills but did not define these skills. The present research broke this down into the ability to solve physics problems and the ability to explain physics concepts. It established that each year about half of the teachers who had assessed against the Level 2 PUS agreed that they were appropriate for assessing students' ability to solve physics problems but about a third of teachers disagreed that they were suitable for this purpose (§4.1.3b). The latter commented that the PUS did not enable assessment of higher order problem solving that required a synthesis of ideas. Some teachers expressed concern that the rigid prescribed format of problem solving dictated by the performance criteria sometimes prevented students who obviously understood the problem from attaining credit. In addition, teachers were uncertain about the appropriateness of the Level 2 PUS for assessing students' ability to explain physics concepts and doubted that the divergent and creative thinking involved in explaining concepts could be adequately assessed by the PUS (§4.1.3c).

The reason given for these concerns was that the performance criteria did not address higher level skills and placed too much emphasis on presentation skills such as the requirement to answer in a sentence format, the insistence on a certain number of significant figures and SI units (§4.3.3.1). These concerns are partly addressed by recommendation 7.3 that suggested broadening the performance criteria. In addition the criteria need to be reconfigured to reduce the over emphasis on presentation skills and address higher-level skills. Commenting on US in general, Coogan (1996:94) suggested the inclusion of criteria that address "planning, synthesis and evaluation". This suggestion is appropriate for the Level 2 PUS. The suggestion related to planning is addressed by recommendation 7.2 since PUS 6384 prescribes an extended student investigation that includes specific references to planning. The inclusion of performance criteria that address synthesis and evaluation would improve the construct validity of the PUS.

Recommendation 7.5

The performance criteria should be reconfigured to reduce the emphasis on presentation skills and address higher level skills such as synthesis and evaluation.

A synthesis of the findings related to curriculum fidelity indicates that the curriculum representation of the PUS was high, that the credit and level allocations were appropriate, that they enabled a high level of curriculum coverage and were generally suitable for assessing practical work and straightforward problem solving skills. However doubts remain about the suitability of the PUS for assessing higher order problem solving skills and the ability to explain physics concepts. Overall it can be concluded that assessment against the PUS has satisfactory curriculum fidelity but that this can be further enhanced by the implementation of recommendations 7.1-7.5.

7.1.2 Concurrent validity

The research established that assessment against the Level 2 PUS had high concurrent validity with assessment for SFC physics. This was shown in 1996 and 1997 by the high correlation between the total number of Level 2 PUS credits achieved by students and their end of course SFC physics grades (§4.2). In 1996 there was some teacher speculation that the correlation might be higher for boys than for girls because the latter were thought to show more attention to the details specified in the performance criteria. This was investigated in subsequent years but was not found to be significant. The high correlation between SFC and PUS indicates that the PUS produces a similar distribution of results to SFC and is a potentially valid replacement for SFC. Furthermore the PUS credit distribution has the advantage that it is generated without reference to the SC mark distribution and avoids the problems associated with scaling.

7.1.3 Validity of reporting student achievement

Each year at least half of the teachers who had assessed against the PUS felt that credits were valid for indicating student achievement in Year 12 physics. A comparison with SFC showed that teachers rated credits equally valid as SFC grades (§4.3.1). Teachers liked the diagnostic value of the skill profile that is established by the student record of the elements that have been or are yet to be achieved and the clear link between the learning outcomes and students' records of achievement. Students supported the concepts of explicit learning outcomes, credit at the unit level and lower stakes assessment close to learning and felt that this facilitated thorough learning (§4.4.1). This indicates that while the PUS are used primarily for assessment for certification, they also have diagnostic, ipsative and formative functions that contribute to the teaching and learning process.

Teachers felt that the competent/not yet competent split for indicating student achievement was insufficiently representative for reporting on students' achievement. They were concerned that assessment against the PUS did not provide a mechanism for reporting on partial achievement and excellence. This echoes concern raised by Austin (1996), NZIP (1996), NZPPTA (1997a), Patterson (1996), Rawson (1997) and Smithers (1997). Over three-quarters of the teachers surveyed wanted criteria for recognising and rewarding excellence to be built into the PUS (§4.3.4). Analysis of the students' questionnaires revealed that students mirrored teachers' concerns that the competent/not yet competent split for reporting achievement was insufficient and that they wanted recognition of different levels of achievement including excellence to be built into the PUS. In addition, lower ability students commented that the level of achievement required for credit was too high for some PUS (§4.4.1). The implementation of the following recommendation would address the concerns raised by students and teachers.

Recommendation 7.6

Criteria for the recognition of different levels of achievement including excellence should be built into the PUS. The level of achievement required for credit should be attainable by a majority of students.

Teachers expressed concern that the compartmentalised, discrete credit structure did not reflect the holistic nature of physics because it encouraged students to achieve the curriculum aims piecemeal. In an attempt to ensure that students reach the required standard, some teachers taught directly to the US and felt that this restricted the scope of their programme by preventing them from adopting a more contextual approach to the curriculum. This reflects the concern raised by Smithers (1997) that the essence of subjects may be adversely affected by trying to express them as numerous performance criteria.

In support of SFC, teachers commented that the grades provide an efficient mechanism for ranking students, rewarding excellence and differentiating between different levels of achievement. These aspects are currently lacking in assessment against the PUS but the introduction of broader performance criteria and recognition of different levels of achievement suggested by recommendations 7.3 and 7.5 respectively may address these concerns to some extent.

The process by which credit was awarded was generally transparent to students who had been assessed against the Level 2 PUS. The reason for this was that students generally had the process explained well by their teachers and had received an NZQA leaflet. Students had a similar understanding of the reporting processes for the PUS and SFC but students in large schools understood the process by which credit is awarded for the Level 2 PUS less well than students in medium and small schools (§4.3.2). A possible explanation for this is that smaller schools have fewer physics classes and teachers and consequently less emphasis on inter-class moderation that is generally not transparent to students. In addition smaller class sizes allow teachers more time to explain the credit allocation process to individual students.

There were four aspects of the reporting process that were poorly understood by students. These included uncertainty about what qualification they were working towards, uncertainty about the credit weighting of each Level 2 PUS, and how many credits are required for the National Certificate. Students were also unsure about US requirements for entry into Year 13 or tertiary study.

Recommendation 7.7

The structure of the National Certificate in Educational Achievement needs to be fully explained to students. The course descriptions provided by schools and tertiary institutions need to incorporate detailed references to the Framework and US requirements for entry into Year 11, 12 and 13 as well as explaining the pre-requisites for entry into tertiary study.

The *Assessment Guide: Physics* outlined a standard model for recording the results of assessment against the PUS on students' scripts. It suggested the use of specially designed answer sheets that included tick boxes to indicate the performance criteria for which students had met the competence requirements outlined in the judgement statements in the assessment schedule. Each year a substantial majority of teachers found this system useful for describing students' achievement (§4.3.3.1). They felt that it provided useful diagnostic information on what skills each individual student had mastered or not. Teachers found this information helpful in targeting assistance at both the individual student and class levels.

At the cluster meetings held during the trial in 1996, teachers were instructed not to provide any additional written feedback on student scripts or specific guidance on where mistakes had been made. This system was adopted because students were given the opportunity to resubmit their scripts after being given the opportunity to correct minor mistakes. This practice restricted teachers from giving specific feedback on student work because it would give students who resubmitted work an unfair advantage and therefore could adversely affect the validity of their results.

This lack of feedback was not popular with students because it led to uncertainty about the usefulness of the results of assessment against the PUS for describing achievement (§4.3.3.2) or for describing how achievement could be improved (§4.3.5). Students found SFC assessment results significantly more useful for both of these purposes. They felt that ticks did not provide sufficient feedback and preferred more detailed written feedback that identified exactly where they went wrong. This may be a reason for the finding by Fitzsimmons (1997) that most students preferred exams rather than US and that many students wanted percentage marks or achievement grades rather than the pass/fail nature of US. This concern is partly addressed by the implementation of recommendation 7.6 which suggests including additional levels of achievement for reporting students' achievement.

Students from large schools found the results of assessment against the Level 2 PUS less useful for describing how well they performed on a task than students from medium and small schools. A possible explanation for this is that smaller class sizes allow teachers to spend more time on re-submissions and provide more reassessment opportunities.

This raises the concern that there may be variations between providers and is supported by the finding that in 1998, 35% of the moderators felt that conditions allowing re-submissions varied between schools and between students and there were no definitive guidelines about what constitutes ground for resubmission (§5.7).

Since the practice of resubmission impacted negatively on the quality of the feedback teachers were able to give on students' work and may adversely affect comparability between schools, it should be abandoned.

Recommendation 7.8

The practice of resubmission[4] should be abolished.

The implementation of this recommendation would allow teachers to give more detailed diagnostic feedback on student' scripts and has the additional

[4] Resubmission involves providing students with the opportunity to make changes to their original script and have it remarked. It is different from reassessment, which involves additional assessment using a different activity to assess the same topic at a later date.

benefit of contributing to improving comparability between schools by removing a source of variability in assessment conditions.

7.1.4 Consequential validity

Aspects of consequential validity investigated by the research included students' satisfaction with the way they were assessed against the PUS and its impact on students' enjoyment of the Year 12 physics course, learning, and motivation. The research also addressed the impact of assessment against the PUS on classroom teaching and teachers' enthusiasm.

Over the period of the study there was a significant increase in students' level of satisfaction with the way they were assessed against the PUS but students were significantly more satisfied with assessment for SFC than assessment against the PUS. Students from large schools were significantly less satisfied with the way they were assessed than students from medium and small schools (§4.4). The latter finding may be due to the reduced amount of time teachers are able to allocate to resubmission and reassessment of individual students.

There was a number of aspects of assessment against the PUS that students liked. They valued the clearly specified learning outcomes that are stated in each element and liked lower stakes assessment close to learning and credit at the unit level. They supported the concept of reassessment and felt that this facilitated thorough learning. In comparison with SFC they liked the absence of scaling and predetermined credit distributions and the fact that each subject and student is treated independently (§4.4a).

Students did not like dual assessment and the resulting over assessment. (§4.4.1, §6.1.3). They were uncertain whether the PUS credits were recognised by employers and for entry into tertiary study. They did not like the fact that assessment against the PUS did not differentiate between different levels of achievement. They felt that the all or nothing nature of the assessment did not recognise partial achievement and that it was difficult for less able students to get any credit. More able students complained about the lack of recognition of excellence and competition (§4.4b). These concerns were addressed by recommendation 7.6.

Overall however, each year the majority of students who were dually assessed against the Level 2 PUS and SFC found the Year 12 physics course enjoyable and they enjoyed the course similarly to students who were not assessed against the PUS (§4.4.4). Student enjoyment of the course increased significantly between 1996 and 1998 and between 1997 and 1998 (§4.4.4). This may be explained by the introduction of the new physics curriculum and the revised prescriptions in 1998 which were associated with an increase in emphasis on contextual teaching and extended investigations. These findings do not support Batchelors' (1996) claim that students hate the PUS.

The literature review established that teachers considered assessment against the PUS to have a negative effect on the learning of more able (NZIP, 1996) and less able (NZIP, 1996; NZPPTA, 1997a and Vlaardingerbroek, 1996) students. Furthermore the NZIP (1996) survey found that a majority of the teachers surveyed felt that assessment against the PUS did not promote excellence and industry.

The present research found that in 1996, 44% of teachers were uncertain about the impact of assessment against the PUS on student learning in general, mainly because they felt that the impact was different for different students. This was followed up in 1997 and 1998. The results showed that over that period, about half (50%-57%) of the teachers felt that it impacted positively on the learning of average students but approximately half felt that it had a negative impact on the learning of less able students (§4.4.2). This finding concurs with that of the NZQA trial school survey (NZQA, 1996).

In 1996 teachers were unsure about the impact of assessment against the PUS on students' motivation. In subsequent years about 45% of teachers felt that it had a positive impact on students' motivation and a minority felt it impacted negatively (§4.4.3). Teachers commented that assessment against the PUS had the biggest impact on the motivation of average students because they work harder to complete a US if they have already got a couple of elements. Furthermore the awareness of reassessment has focused them on the learning objectives. It was felt that assessment against US had a negative impact on the motivation of less able students because they get no record of success and consequently give up.

Teachers felt that assessment against the PUS did not act as a motivator for above average students. Despite this and the earlier finding about the lack of recognition of excellence, approximately half (48%-61%) of teachers felt that the assessment against the PUS had a positive impact on the learning of more able students. This was opposite to the NZQA trial school survey finding. It may be that since students were dually assessed they viewed the two types of assessment as complementary. They may have liked the diagnostic nature of the PUS and the competitive aspects of assessment for SFC.

Over the period of the study, 39-50% of teachers surveyed felt that assessment against the PUS impacted positively on their classroom teaching and about a third felt that it impacted negatively (§4.4.6). In 1996 and 1998 about half of the teachers who assessed against the PUS felt enthusiastic about their involvement and approximately a quarter were unenthusiastic (§4.4.5). In 1997 teachers felt less enthusiastic. This may be due to the unsettling effects of industrial action and political uncertainty.

Teachers identified a number of aspects of assessing against US that impacted positively on curriculum delivery and classroom practice (§4.4.6a).

They liked having clear objectives to assess against and that the assessment results are diagnostic and identify what a student can and cannot do. They felt that the implementation of US assisted the professional. development of teachers through the three-day training programme and the interaction with the moderator and other teachers. This was seen as especially valuable in small and or geographically isolated areas.

Assessment against the PUS was seen to assist programme evaluation by providing teachers with a class profile of results that could be used diagnostically to identify aspects for targeting teaching programmes and performance. These aspects were previously masked by global percentage student results in tests for SFC. Assessment against US is a form of diagnostic assessment that can identify specific areas of the course with which an individual student is having difficulty. It is also ipsative because it helps individual students by providing guidelines for future learning.

Assessment against US was seen by teachers to validate school-originated courses because the credits could be registered on the National Qualifications Framework. This enabled greater flexibility in course design. Some schools assessed against level 1, 2 and 3 in a Year 12 Course, whereas others had introduced electro-technology units in their physics programme. These vocational US contribute towards qualifications that can be completed in tertiary study or the workplace. Before the implementation of the Framework, local school courses did not have any national recognition.

Teachers identified a number of disadvantages of assessing against the PUS (§4.4.6b). They considered the timeline for implementation to be unrealistic. This coupled with the increase in the workload associated with the implementation of US at a time when other demands such as the implementation of the new curriculum have also increased is affecting the morale of some teachers adversely. Some teachers commented that they withdrew from assessment against the PUS because they became disillusioned with the administrative demands of moderation, recording, portfolio keeping, filing and reporting. This was seen to be at the expense of teaching and the development of new strategies. These issues and associated recommendations are discussed in detail in the section on manageability (§7.3).

Findings related to consequential validity were that students were more satisfied with assessment for SFC than the PUS although satisfaction with the latter increased significantly over the period of the study. Assessment against the PUS was seen to have a positive impact on the learning of average and more able students but a negative impact on the learning of less able students. This may be because the level for credit was set too high and was addressed by recommendation 7.6. A majority of students who were dually assessed enjoyed the Year 12 physics course.

More teachers felt that assessment against the PUS impacted positively on their classroom teaching than negatively. Likewise more teachers were enthusiastic about their involvement in assessment against the PUS than were unenthusiastic. Subject to addressing the manageability concerns raised by teachers, it may be concluded that assessment against the PUS had satisfactory consequential validity.

The research established that assessment against the PUS had satisfactory curriculum fidelity and concurrent validity with SFC and that credits were valid for indicating students' achievement. Overall consequential validity was satisfactory. It can be concluded therefore that assessment against the Level 2 PUS is a potentially valid replacement for SFC. Teachers did however identify a number of areas in which validity could be enhanced and these were addressed by recommendations 7.1-7.8.

7.2 Moderation of assessment against the PUS

Smithers (1997) felt that the moderation system's aim of achieving comparability of assessment against Unit Standards across nearly 450 secondary schools was dauntingly difficult to achieve. A number of commentators expressed doubts about the consistency of the moderation process (Croft, 1993; Irwin, Elley and Hall, 1995; Mann, 1997; Morris, 1996 and NZPPTA, 1997).

The present research confirmed that there was a number of potential threats to achieving comparability of assessment standards across schools. The threats identified by moderators included assessment and qualification policy uncertainty, untrained providers, insufficient sampling, reduced standards for reassessment and concerns about the authenticity of student work and sufficiency of evidence. In addition they identified possible variations in the interpretation of the PUS, school-based assessment conditions, internal moderation procedures, grounds for resubmission and moderator and end-point assessor judgements (§ 5.7).

The research addressed these concerns with a detailed investigation into all aspects of the moderation process. These included internal moderation, the moderation of assessment plans, assessment activities and assessor judgements, the NPA, check moderation and communication within the moderation system. This section discusses the research findings related to each of these aspects and makes research-based recommendations for improving the moderation system.

The research showed that schools generally had similar well-defined internal moderation policies that were set out in their accreditation documents (§ 5.3.1). The analysis of the1998 moderator questionnaire showed that 65% of the moderators were either "Satisfied" or "Very satisfied" that the conditions under which assessment occurred were comparable between schools.

Moderators did express concern that there was no check on internal moderation procedures between different classes in the same school (§5.7). This supports Rawson's (1997) concern that internal moderation is outside the jurisdiction of the MAP and therefore leads to uncertainty about the comparability between different assessors within the same institution. This concern could be addressed by asking providers to include a copy of their internal moderation policy with the assessment plan that is submitted early in the year.

Recommendation 7.9
The providers' internal moderation procedures should be submitted for approval at the start of each year.

The induction of new moderators is an important first step towards achieving comparability. Moderators felt that the moderator training programmes were highly effective in enabling them to develop an understanding of the moderator role, the physics MAP and the moderation of assessment activities and schedules. Several moderators commented that the programme should be extended by one day to provide more practice moderation of assessment activities and schedules (§5.2).

Recommendation 7.10
The length of initial moderator training should be increased to three days and a larger portion of the training programme should be devoted to practice moderation of assessment activities.

Each year a substantial majority of moderators considered the moderation of assessment plans to make a satisfactory contribution to achieving comparability because it enabled them to target the same PUS for different providers. In addition it was seen to help moderators plan their workload and encouraged providers to plan their assessment programme earlier in the year and prepare assessment activities and marking schedules ahead of time. This was seen to have a positive effect on validity and presentation (§5.3.2).

The moderation of assessment activities is the main component of the MAP. Each year a clear majority of moderators felt that this aspect of the MAP made a satisfactory contribution to achieving comparability between schools. Moderators felt that comparability was enhanced by the publication of the *Assessment Guide: Physics*, the NPA's, the CD of assessment activities and pre-moderated activities because these were used by a large number of providers and provided models for providers to design their own assessment activities. In 1996 and 1997 a majority of teachers agreed with moderators but in 1998 teachers were significantly less certain. This may be as a result of the political uncertainty about the future of the PUS.

The national approval ratio of assessment activities that were submitted for the first time to moderators is another indicator of comparability. This ratio increased significantly from 52% in 1996 to 70% in 1998. This indicates an increase in the ability of teachers to design activities that met the requirements of the standard and hence comparability between schools (§5.5b).

In 1997 and 1998 a large majority of moderators felt that the consistency between moderators was satisfactory (§5.5a). This showed that Carter's (1996) concern about the lack of consistency between Local Moderators was largely unfounded in the case of the Level 2 PUS. The annual moderator agreement trials provided further evidence to support this and showed that over the period of the study there was a high and growing level (70%-80%-84%) of consistency between moderators when moderating a common assessment activity. Over the same period the mean and standard deviation of the time it took to moderate a similar activity decreased significantly (§5.5b).

Further evidence for increasing consistency came from the investigation into the check moderation carried out by the Regional and National Moderators. In 1997 and 1998, a majority of moderators felt that the check moderation process was satisfactory in achieving comparability between providers. In 1998 check moderation was seen as less important because the increasing confidence of moderators and the increased expertise of providers meant that the check moderation process picked up very few inconsistencies (§5.3.5).

Each year, a majority of moderators and teachers felt that the verification of assessor judgements also made a satisfactory contribution to achieving comparability. This finding allays Salter and Hayden's (1996) concern that because there is a large number of moderators and providers there is a potential for inconsistencies in assessor judgements to develop. Moderators felt confident that teachers were generally very accurate in their marking and recommended a reduction in the number of items of student work that need to be submitted for verification. Over the period of the study teachers became significantly less certain about the comparability of assessor judgements and said they had no basis for judging this. They also felt that some providers could be more lenient than others (§5.3.4). The end-point assessor judgement agreement trials carried out in 1996 and 1997 provided a more objective indicator of consistency and showed that the mean percentage of agreement of end-point assessor judgements was consistently high (§5.6). In 1998 the number of items of student work that have to be submitted to the Local Moderator was reduced from six to four. Given the high level of end-point assessor judgement agreement, this number could be further reduced to three and should include a pass, fail and partially complete element or PUS.

Recommendation 7.11

Reduce the number of items of student work that need to be submitted to the Local Moderator for the verification of assessor judgements to three.

Chamberlain (1996) felt that there was pressure on teachers to be less rigorous with reassessments than the original assessment activity because the MAP does not cover it. Moderators also identified reassessment as a threat to comparability. There is variation between schools about the number of reassessment opportunities offered to students. A national guideline of one opportunity per PUS would introduce consistency between schools and also reduce the workload for teachers. Reassessment is aimed directly at students who are on the competency borderline. It can be argued that they determine the level at which students are credited with a PUS more definitively than the original assessment activity. It is important therefore that reassessment activities should be moderated. It would be convenient if a parallel reassessment activity was submitted at the same time as the original assessment activity.

Recommendation 7.12
Reassessment should be limited to one opportunity and reassessment activities should be moderated.

In 1996 and 1997 the MAP for physics included an NPA. In 1997 a substantial proportion of the physics moderators felt that the NPA was necessary to the moderation process but a majority of teachers were unsure of its organisation or felt that it was unsatisfactory (§5.4). The reasons for this were that it was not compulsory, and that the NZQA specified timing of the NPA did not relate to a logical time in a provider's programme. Forty-three percent of the moderators felt that it would be more valuable if the NPA's could be held at any time during the year at a naturally occurring time in a provider's programme.

Teachers felt that the NPA should not be a reference test to be taught to as a focus for the year and that the moderation system had enough checks on consistency built into it without the NPA. Teachers commented that NZQA didn't use the results to check national consistency and in 1997, a majority of the moderators disagreed that the NPA results from each school should be used as a statistical check on the way schools award credit. The NPA was abolished in 1998 and replaced by pre-moderated activities. These were seen as a way of retaining the positive aspects of the NPA while at the same time addressing its criticisms. In 1998, 95% of moderators supported the introduction of the pre-moderated activities (§5.4b).

Recommendation 7.13

The NZQA should provide more pre-moderated activities and upgrade the assessment guide to an item bank of pre-moderated assessment activities.

This recommendation, coupled with allowing providers to quote reference in their assessment plan to pre-moderated activities would reduce the external moderation requirement and reduce the workload for teachers and moderators.

The effectiveness of communication within the moderation system plays an important role in achieving comparability. Each year a large majority of moderators judged the communication with the National Moderator to be very satisfactory. They particularly valued the National Moderator's newsletters but felt that the frequency of these should be increased. They also felt that it would be valuable to publish a regular provider newsletter to update teachers on PUS interpretations (§5.3.6a). These could be published on the NZQA web site.

Recommendation 7.14

The National Moderator should publish a newsletter for moderators once a term and a newsletter for providers biannually. These could be published on the NZQA web site.

Each year a large majority of Local Moderators found their communication with their Regional Moderator to be satisfactory. In addition, both Local Moderators and Year 12 physics teachers felt that the communication between Local Moderators and providers was satisfactory.

A special feature of communication within the moderation system is the cycle of national, regional and local meetings. In 1998 moderators felt that the regional meetings were satisfactory but that the value of these meetings is increased if moderators are asked to moderate a common assessment activity in preparation for discussion at the meeting. In order to avoid regional variations it is important that consensus on moderation issues is reached at the regional meetings. Moderators felt that the presence of the National Moderator at all of the regional meetings helped to establish national consistency.

Recommendation 7.15

To improve the consistency of the moderation process, moderators should be asked to moderate a standard assessment activity in preparation for the regional meetings.

In contrast moderators felt that the meetings of Local Moderators with providers were not satisfactory. Moderator opinion was divided on whether these meetings were necessary. Since attendance at the meetings was not compulsory and they were not funded by NZQA, attendance was poor. In addition the wide geographical spread of providers in some clusters made the organisation of the meetings difficult. The local provider meetings were not seen as valuable as the funded cluster meetings NZQA ran during the 1996 trial. These meetings were found useful by a very high proportion of teachers. They particularly valued the opportunity to discuss issues with colleagues and were appreciative of the NZQA support people who ran the meetings.

Recommendation 7.16

Organisation and attendance at provider meetings could be improved by reallocating providers to moderators to form clusters of local schools. The meetings should be fully funded by NZQA and take place early in the school year.

This would enable moderators to meet the provider contacts early in the school year to outline moderation requirements and discuss issues and provider questions.

The research established that the moderator training, the moderation of the assessment plan, assessment activities and assessor judgements, and communication within the moderation system all contributed satisfactorily towards achieving comparability between providers. In addition the annual moderator and end-point assessor judgement agreement trials indicated high levels of consistency between moderators and comparability between providers. Overall it can be concluded that the moderation system has satisfactorily established and maintained comparability between providers and that this can be further enhanced by the implementation of recommendations 7.9-7.16.

7.3 The manageability of assessment against the PUS

The discussion so far has established that assessment against the Level 2 PUS has satisfactory validity and comparability between providers. It remains to establish whether it is manageable for teachers, students and moderators. This is a crucial component of the quality management of assessment because the implementation of a new system of assessment that is highly valid and reliable but not manageable may be opposed or even halted by stakeholders. This scenario was illustrated the 1996 PPTA freeze on assessment against US. The literature review established the manageability of the workload associated with assessment against US as a major concern for teachers (NZPPTA, 1997a; Rawson, 1997; Rosser, 1996). More specifically 66% of the teachers surveyed by the NZIP (1996) felt that the workload was unsustainable. The current research found that in 1996 and 1997 the workload associated with the implementation of assessment against the PUS was higher than that for the established SFC. This was not surprising given the fact that teachers had built up assessment resources for the latter over a longer period. Teachers identified participation in professional development, assessment task design, meeting moderation requirements, marking, record keeping, resubmission, reassessment and dual assessment as aspects of assessment against the PUS that contributed to their workload. By 1998 teachers had become more familiar with the new system of assessment and had developed their own assessment resources. In addition, they were provided by NZQA with additional resources such as the Assessment Guide updates, pre-moderated activities and a CD item bank of assessment exemplars. Consequently the workload had settled down and there was no significant difference in the workloads associated with the two systems of assessment (§6.6.1).

Despite this finding, concerns about perceived workload and excessive administrative demands remained. Recommendations to ease this situation need to address these manageability concerns while ensuring that assessment against the Level 2 PUS retains optimum validity and reliability.

A number of the recommendations already made in relation to validity and reliability have the additional advantage of making assessment against the Level 2 PUS more manageable for teachers. These included the suggestions to abolish the practice of resubmission (recommendation 7.7) and to limit the number of reassessment events to one (recommendation 7.10). In addition the suggested provision of an item bank of pre-moderated activities (recommendation 7.11) and the reduction in the number of items of student work that need to be submitted for moderation to three (recommendation 7.7) will have the effect of reducing the workload for teachers.

There are other gains in manageability that can be made. Teachers identified dual assessment for both the PUS and SFC as a major contributor to the workload. While dual assessment was necessary for the schools that were involved in the 1996 trial of the PUS, it contributed unnecessarily to the workload in subsequent years. The present dual assessment situation is a consequence of the political uncertainty associated with the hiatus in New Zealand's qualification policy (§1.3). A resolution of this situation would enable teachers to concentrate on one form of assessment only and greatly reduce their workload. Some schools had anticipated this situation and abolished assessment for SFC.

Recommendation 7.17
The Framework implementation timetable should avoid prolonged periods of dual assessment.

There are at present 12 PUS at Level 2 of the NQF. Many of these are content-based but repeat elements and performance criteria that address similar skills. A minority of teachers expressed concern about the large number of Level 2 PUS. Students who complained about over assessment supported this view. There is a considerable overlap between some PUS and the same performance criteria are repeated in different PUS. A reduction of this overlap would not affect validity but make assessment against the PUS more manageable by reducing the number of assessment events. This could be achieved by reducing the total number of PUS at Level 2 to 6-8. These standards could be skills-based and include special notes that allow teachers to assess in different content areas.

Recommendation 7.18
The number of Level 2 PUS should be reduced from 12 to 6-8 and overlap between the PUS should be kept to a minimum.

The original PUS used complex range statements. Range statements prescribe the content areas in which each element must be assessed and were included to ensure curriculum coverage. This meant that the element and associated performance criteria had to be met for each aspect of the range statement. For an element that had four items of content in its range statement and three performance criteria[5], teachers had to assess each performance criterion four times. This meant that teachers had to record 12 separate aspects of students' achievement for just one element. This approach was very prescriptive and did not allow assessors to design assessment activities that sampled the content in the same way as traditional assessment for SFC. It was also very time consuming and led to over assessment and complex systems for recording students' achievement. This contributed to teachers' complaints about excessive administration and pressure on teaching time (§4.4.6) and students' complaints about over assessment, the pedantic nature of the performance criteria and the difficulty of achieving credit because of all or nothing reporting. (§4.4.1) A reduction in the range statements would address these concerns but could adversely affect validity. This concern could be addressed by transferring the content from the range statements to the special notes.

Recommendation 7.19

The number of range statements in Level 2 PUS should be reduced. The content could be transferred to the special notes that accompany each PUS.

The literature review indicated that teachers supported calls for time allowances (Rawson, 1997; Batchelor, 1996) and administrative support (Rawson, 1997). Implementation of the following recommendations would address these concerns.

[5] See element 2 of PUS 6379.

Recommendation 7.20

Providers that assess against the PUS should be given a time allowance or administrative support.

For the moderation process to be sustainable it needs to be manageable for moderators. In 1996 all of the moderators were new to the role and had to familiarise themselves with the MAP both as a moderator and a provider in their own schools. Consequently a majority of moderators found the workload difficult to handle. This situation improved in subsequent years. By 1998 the majority of moderators described their moderation workload as average. The pressures on moderators were not due to the quantity of work but related to the pressure of meeting deadlines and the intermittent nature of the work. To alleviate the workload, moderators recommended that the number of providers should be limited to 8 for those moderators who were teaching full time (§6.1.2).

Recommendation 7.21

The number of providers allocated to each moderator should not exceed 8.

While all the moderators used the standard forms as a summary of their comments, the forms were inadequate for the task. Some moderators wrote letters to providers to accompany the forms. Providers complained about having to repeat information on the forms such as contact details etc. which are already on file and felt that the moderation forms could be simplified and reduced in number.

Recommendation 7.22

The NZQA moderation pro-formas should be redesigned to make them more user friendly for providers and moderators.

The main effect of US assessment on student workload and the number of assessments is the extra workload due to dual assessment, an increase in the frequency of assessment, a reduction in the length of assessments, a greater variety of assessments and the opportunities offered by reassessment. Apart from the dual assessment issue these features of competency-based assessment were generally viewed positively by students.

Students felt that the workload in their Year 12 Physics course was similar to that of other subjects. Students who were assessed against the Level 2 PUS described the workload in their Year 12 Physics course as slightly but significantly less than students who were assessed for SFC only. A possible explanation is the difference in the nature of the assessment used. Assessment for SFC typically takes a full class period and occurs at the end of each topic. Since there are no reassessment opportunities the individual summative assessments carry a large emphasis. Schools that assessed against the PUS typically had a larger number of smaller formative assessments that were carried out close to learning and were often integrated into ordinary lessons. Since students were offered reassessment opportunities the emphasis on individual assessments is less than for SFC. It is possible that students perceive these assessments as part of the learning process and do not regard them as contributing to an increase in workload. The workload in Year 12 physics courses increased significantly between 1996 and 1997 but there was no significant difference between 1996 and 1998. A possible reason for this is the increase in the number of schools that practised dual assessment for both US and SFC in 1997 (50 in 1996 and 65 in 1997[6])

Each year of the study the number of assessments for physics was about the same as for other subjects for both schools that assessed against the PUS and schools that assessed for SFC only. The number of assessments for students who were dually assessed against the PUS and SFC was greater

[6] 1998 data not available.

than the number of assessments for students who were assessed for SFC only.

Students who were dually assessed against the PUS and SFC referred to the increased frequency but smaller size of PUS assessments. Student felt that since the assessments were closer to learning and reassessment opportunities were offered, the assessments were more valid and seen to be fairer.

Another aspect of manageability is the adequacy of resources and the support provided by the NZQA for assessment against PUS. A majority of teachers felt that NZQA should have supplied them with a personal paper copy of the PUS. They found the CD Framework Explorer difficult to use and did not always have access to a computer or the NZQA web site.

Teachers felt that the *Assessment Guide: Physics* was useful as an item bank of assessment activities but they expressed concern at the number of mistakes in the guide. They wanted the activities in the guide to be pre-moderated. This suggestion has been addressed by recommendation 7.11. Teachers found the guide useful for suggestions on record keeping and implementation tips.

The three-day training programme in competency-based assessment was found useful by a high proportion of teachers but some teachers expressed concern that it was not available in all regions or to new staff and teachers who started assessing against the PUS in 1997 and 1998.

Recommendation 7.23
All providers that assess against the PUS should receive NZQA funded training.

Since the trials and training group within NZQA has been disbanded, additional training could be contracted out to Advisory Services.

Following the implementation period, the workload associated with assessment against the Level 2 PUS was similar to that for SFC. Teachers' complaints about manageability referred mainly to the practice of dual assessment and administrative demands. Overall it may be concluded that assessment against the Level 2 PUS is generally manageable for teachers and that it may be further enhanced by the implementation of recommendations 7.17-7.23.

7.4 Conclusion

Assessment against the Level 2 PUS is generally valid. There are unresolved issues related to the atomisation and micro definition of learning outcomes, the suitability of the PUS for assessing higher level skills and conceptual learning. Teachers liked the diagnostic value of the reporting process but would like to see it address the recognition of different levels of achievement including excellence.

Overall, the Physics MAP achieved a high level of consistency of moderator decisions and comparability of assessment standards between schools. An investigation into the components of the moderation system established that the moderation of the assessment plan, assessment activities and schedules, assessor judgements check moderation and communication within the moderation system contributed effectively towards achieving comparability between schools. The annual moderator agreement trials and end-point assessor judgement-agreement trials provided additional evidence for this. Moderators identified a number of potential threats to achieving comparability between schools. These included insufficient provider training, high provider and moderator workload, and variability in internal moderation, PUS interpretation and conditions for resubmission and reassessment. They also identified light sampling and authenticity of student work as potential threats.

NZQA needs to be careful that future changes to the moderation system do not adversely affect comparability between providers.

Assessment against the Level 2 PUS was initially difficult to manage for teachers and moderators but improved over the period of the study. Dual assessment and administrative demands were identified as major areas to be addressed. The resources and training supplied by NZQA were generally adequate. Teachers liked the provision of the *Assessment Guide: Physics* but wanted funded support for attending local provider meetings and training opportunities for teachers who had missed out on the initial training rounds and additional funding for Advisory Services during the implementation period.

Since assessment against the Level 2 PUS had satisfactory validity and comparability and was generally manageable for stakeholders it may be concluded that it is an effective potential replacement for SFC. However the recommendations highlight a number of areas that need to be addressed. Chief among these are the incorporation of higher level skills, recognition of different levels of achievement including excellence, reduction in the amount of assessment, elimination of dual assessment and the incorporation of external assessment. A more inclusive NQF that allows for a greater variety of assessment approaches would enable greater freedom in the selection of the most suitable approach for a particular subject and assessment context.

7.5 Directions for future research

This research could be replicated in different curriculum areas in order to get a broader overview of the validity, reliability and manageability of competency-based assessment.

There needs to be some follow up research to investigate different approaches to incorporating the recognition of excellence and different levels of achievement into competency-based assessment against US.

The question of whether it is possible to assess higher level skills and conceptual learning using a competency-based approach has not been fully answered. Teachers' concerns may simply be related to the present format of the PUS and a revision could include a greater emphasis on higher level skills. It may be however that it is not possible to assess higher level skills because it is not possible to write performance criteria that use behavioural terms to describe mental processes. This issue needs to be further investigated.

The moderator agreement trials investigated the consistency between moderators on a common activity. This was somewhat artificial since in reality providers submit a range of different activities for the same PUS. An alternative research strategy would be to collect a national sample of different assessment activities for the same PUS and moderate these using a consensus panel approach.

The present research can be used as a baseline to evaluate the effectiveness of the implementation of any of the recommendations and any subsequent changes to Year 12 physics assessment such as the introduction of Achievement Standards in 2001.

Chapter 8

Implications for the NQF

The present study consisted of a detailed longitudinal evaluation of the quality of assessment against the level 2 PUS. The research findings and issues raised by the participants in the study have wider relevance beyond this immediate context. This final chapter synthesises the findings of the present study and the research literature to discuss the implications of national implementation of assessment against US for all conventional school subjects at Levels 1-3 of the NQF.

The discussion commences by addressing two fundamental systemic issues. These are whether assessment against US is suitable as a replacement for the current system of senior secondary school qualifications and whether it can achieve an acceptable level of public credibility and national consistency. The discussion then proceeds to consider the likely consequences of national implementation for students, teachers and schools in general. The conclusion summarises the key issues that need to be resolved before the system is fully implemented at a national level and makes associated recommendations for improving the system.

8.1 Is assessment against US a suitable replacement for the current system of senior secondary school qualifications?

The NQF proposes to replace the current secondary school qualifications with a single model of assessment against US for all subjects in the senior secondary school curriculum. The literature review for the present study identified some fundamental issues that need to be satisfactorily resolved before this system is fully implemented. These issues are encapsulated by the following three key questions.

- Is it possible to write standards that are unambiguous and provide sufficient detail to assess against?
- Is it appropriate to use a single model of assessment against US to assess all subjects in the senior secondary school curriculum?
- What are the consequences of expressing the learning outcomes for each subject in terms of discrete US and elements?

This section discusses these key questions with reference to the research literature and the findings of the present study.

Is it possible to write standards that are unambiguous and provide sufficient detail to assess against?

Coogan (1996: 14) stated that "transparency is a key principle of competency-based assessment". To achieve this US should be specific and unambiguous. Smithers (1997: 78) felt that US could not be stated with the "necessary precision to ensure the fairness, consistency and validity of assessment" and that they should therefore not be used as the common currency for the NQF. This conclusion misses the point that the standards are not defined exclusively by written statements. Saddler (1987) has written about the value of exemplars to clarify standards. Casting the net for the location of the standard more widely, are a number of commentators that argue that US cannot be defined solely by the written standards that are registered by the NZQA but are defined through a process of triangulation between the registered standard, the moderation process and the assessment activities and schedules (Batchelor, 1996; Irwin, 1995 and Irwin, Elley and Hall, 1995).

The debate about whether standards can be expressed precisely and how they are defined, needs to be informed by subject specific investigations. The present study found that the Level 2 PUS had high content validity and were representative of Level 7 of the curriculum. The high level of consistency in moderators' decision making shown by the moderator agreement trials, increasing approval ratios of activities and high level of end-point assessor judgement consistency, indicated that teachers and moderators had a high level of shared understanding of the standards. In the Level 2 physics trial the standards were defined by the registered US, illustrated by exemplars and supported by moderation and professional development. The collective guild knowledge of teachers also contributed to developing a shared understanding of the standards. The NZPPTA Inquiry (1997a) agreed that all of these factors contribute to defining and locating standards.

The implication of these findings for implementing assessment against US in general, is that regardless of where in the assessment system the standard is located, it is possible to develop a high level of shared understanding of standards between teachers. This conclusion is supported by Coogan (1996) and Methven et. al. (1996: 40) who argue that "the difficulty of redefining general education outcomes in standards-based terms seems greatly exaggerated". The overall conclusion is that it is possible to write standards that state educational outcomes but it has to be acknowledged that the full definition of those standards must be supported by effective moderation, professional development and the provision of resources such as exemplars of assessment activities and student work.

Is it appropriate to use a single model of assessment against US to assess all subjects in the senior secondary school curriculum?

Critics of the NQF have argued that assessment against US is valid for vocational and skills-based subjects but that it is unsuitable for assessment of higher order thinking skills that characterise academic subjects. They argue that trying to force assessment for all school subjects into a single model is a fundamental design fault of the NQF (Carter, 1996; Elley, 1994; Salter and Hayden, 1996). This led Irwin (1995:11) to the conclusion that

> The vast and growing range of knowledge and skill cannot fit into the one monolithic bureaucratic framework incorporating one set of levels and one building block.

The present study indicated that for a subject like physics, it is possible to write standards that assess a range of curriculum content at an appropriate level and that the demands of assessment do not impact adversely on curriculum coverage. There is an important difference however between curriculum coverage and depth of learning. The study raised doubts whether higher level skills such as synthesis, evaluation and explanation of concepts can be adequately assessed using a standards-based approach. The study found that the omission of the US approach to assess higher level skills in physics is a consequence of the competency-based approach. The performance criteria were designed to describe the minimum level of achievement required for competence whereas mastery of higher level skills is a feature of the achievement of more able students. Consequently these skills are not included in the criteria for competence. This lead to a mismatch between the assessed domain established by the PUS and the target domain represented by the *Physics in the New Zealand Curriculum*. Since the structure of the US is similar for all school subjects the competency-based approach may be responsible for the failure to address higher level skills in other subjects as well.

Based on this discussion it can be concluded that the US model is probably suitable for vocational courses that lend themselves to being expressed as discrete learning outcomes such as electronics, horticulture and computing. While the model may be suitable for assessing aspects of knowledge-based subjects such as English and history, it may be inadequate for assessing all aspects of the curriculum. This conclusion echoes comments made by other commentators for subjects like mathematics (Neyland, 1994), history (Childs, 1995) and science (Austin, 1996). The approach also appears unsuitable for subjects like art and music that involve performance assessments and require integrated and holistic assessor judgements.

What are the consequences of expressing the learning outcomes for each subject in terms of discrete US and elements?

Another concern about the competency-based approach is that it impacts negatively on classroom programmes by reducing school subjects to collections of US that are described solely in terms of discreet, observable and measurable behaviours (Irwin, 1995; Marshall, 1994 and Salter and Hayden, 1996). Based on his review of the Framework, Smithers (1997:44) expressed the concern that:

> ... the essence of subjects almost seems to disappear in the attempt to express them as numerous performance criteria.

The NZPPTA Framework Inquiry (NZPPTA, 1997a) found that a negative consequence of numerous specific performance criteria and assessment close to learning is that it can start to dominate the design of teaching programmes and lead to the fragmentation of teaching and learning.

The present study found that in the case of the Level 2 PUS there were both positive and negative aspects associated with the nature of the performance criteria. The clearly specified learning outcomes and assessment close to learning were seen by teachers to facilitate diagnostic, formative and ipsative assessment and the evaluation of specific aspects of teaching programmes. However the short-term educational benefits associated with assessment close to learning may have been at the expense of longer term goals such as integration, synthesis and depth of learning.

Physics teachers who participated in the present study perceived the Level 2 physics performance criteria as narrow, pedantic and numerous and found that they did not allow for student answers that followed an unusual or creative approach. A negative consequence of this atomisation of learning outcomes was that it forced a constrained approach to assessment of learning outcomes that were more suitably assessed in an integrated way. While it is possible in theory to design assessment activities that incorporate a number of different PUS, in practice this was difficult to manage. These findings related more to the structure and practice of assessment against US than the specific nature of the physics content and are therefore likely to apply more widely to other subject areas as well.

Based on the discussion of the three specific key questions posed in this section, it is now appropriate to focus on the more general encompassing question of whether assessment against US is a suitable replacement for the current system of senior secondary school qualifications.

The discussion established that it is possible to adequately define standards if the written standards are supported by quality exemplars, professional development for teachers and adequate moderation. It further concluded that a single model of competency-based assessment against US may not be appropriate because it does not allow for the assessment of higher level skills and leads to atomisation of learning outcomes that are better assessed in an integrated holistic context. Furthermore assessment close to learning may not encourage longer-term goals such as depth of learning and synthesis. Based on these conclusions it can be argued that the present model of competency-based assessment against US is not a suitable replacement for the current qualifications.

This does not mean however that standards-based assessment should be abandoned as a mode of assessment for the NQF but indicates that a comprehensive review of the system is necessary. The recommendations made by the present study for improving assessment against the level 2 PUS are similar to those made by the NZPPTA (1997a) for the review of assessment against US in general and provide a useful background for such a review.

In a review of the system, the number of US in each subject area needs to be reduced to enable a more holistic and integrated approach to assessment. This can be achieved by reducing the number of small content-based US and including broader skills-based US that can be assessed in a number of different content areas or integrated across different areas. The accompanying performance criteria need to be broadened in scope and reduced in number. Large unwieldy range statements that prescribe the material that must be assessed need to be made more flexible.

A key recommendation that addresses the need to include assessment of higher level skills is the incorporation of criteria for the recognition of excellence. This is discussed in more detail in section 8.3.

If these recommendations were implemented the reviewed system of standards-based assessment could be a suitable replacement for the current qualifications. The NZPPTA Framework Inquiry (NZPPTA, 1997a: 103) reached the following similar conclusion:

> Assuming that a range of passing grades was available, the Inquiry believes that assessment based on unit standards could be applied profitably to all subjects.

8.2 Is it possible to achieve public credibility and national consistency?

For the new system to gain public credibility there needs to be confidence that there is national consistency and comparability of assessment standards between schools. This in turn depends on the effectiveness of the moderation systems set up for each school subject. Smithers (1997) felt that consistency of assessment against US across nearly 450 secondary schools is too difficult to achieve and other commentators have added their concerns about the effectiveness of moderation (Carter, 1996; Finch, 1994; Irwin, 1995; Mann, 1997; Morris, 1996; NZPPTA, 1997a; Rawson, 1997; Salter and Hayden, 1996). The NZPPTA Framework Inquiry (NZPPTA, 1997a: 68) concluded:

> The literature suggests that the level of consistency that can be achieved through the moderation plans is variable, largely unknown and almost certainly not substantial enough to satisfy the requirement for rigour in the national comparability of standards.

These concerns needed to be investigated by detailed subject specific studies into all aspects of the moderation system. The present study investigated the effectiveness of the Level 2 PUS moderation system in achieving comparability between schools.

The study found that the moderation of assessment plans contributed towards comparability by enabling moderators to view a variety of assessment activities against the same US from a large number of providers. The pre-assessment moderation of assessment activities contributed satisfactorily towards attaining and maintaining comparability between schools. This was supported by the results of the moderator agreement trials for the PUS that indicated a high level of consistency of moderator decisions.

The study showed that the moderation of assessor judgement contributed effectively towards maintaining comparability between schools and the post assessment verification of assessor judgement agreement trials found a high level of consistency between markers using standards-based assessment schedules.

The present study indicated that it is possible to achieve a high level of comparability between schools and consistency between moderators for Level 2 physics and that this improves over time. This finding is similar to the experience of other subjects. The National Moderators for geography (Pepper, 1995) and Chemistry (Torry, 1994) found that initially the quality of assessment activities and teachers' assessor judgements were variable but that these improved over the course of the trials. McIntyre (1995), the National Moderator for mathematics found only minor variations and concluded that national consistency was achieved. The adoption of annual national moderator and end-point assessor judgement agreement trials similar to those piloted in the present study would enable the effectiveness of the quality assurance system at a national level to be evaluated and monitored.

In evaluating the extent to which the moderation of assessment activities, schedules and verification of assessor judgements is achieving comparability between schools, it must be remembered that only 20% of a provider's programme is moderated. This effectively means that there is no check on the quality of 80% of assessment carried out by providers each year. Since providers are told at the beginning of the year which standards will be moderated there is a potential for the quality of the assessment activities for the remaining standards to be lower.

One omission in the moderation plans for all subjects is that reassessment activities are not moderated and consequently there is no check on the quality of these activities. In an environment of competition between schools and expectations by school management and the school community to achieve high pass rates there is subtle pressure on teachers to design reassessment activities that are of a lower standard than the original assessment activity. Since the candidates for reassessment are often near the borderline for credit, reassessment activities are crucial in contributing to the establishment of the standard for competence. In light of this there should be some check on the quality of reassessment activities.

Because the large numbers of moderators in the system poses a potential threat to consistency of moderation, the check moderation process is an important step in establishing consistency during the implementation period. The present study found that in the physics moderation system, lighter sampling might be appropriate once the system is established and a shared understanding of the standards is developed.

The present study found that communication within the physics moderation system and between local moderators and teachers was satisfactory. Teachers valued the professional development associated with the feedback they received from moderators and their involvement in the trials. This has been commented on more widely (Meldrum, 1995; Metcalfe, 1996; Squire, 1996). However since there are limited numbers of secondary advisors and the NZQA funded training is no longer available it is difficult for teachers to obtain professional development when needed.

Because it is possible to achieve high levels of national comparability between schools and consistency of assessor judgements, public credibility of the new system can be achieved. However, if assessment against US was implemented nationally across all senior secondary school levels and subjects there are additional factors that need to be considered. These relate to the sheer size of the system, the number of teachers and moderators involved in the system, the high administrative workload, the variety of deadlines and the operational cost of the system. It appears doubtful whether the current moderation system is affordable and manageable if all secondary schools were involved. The system clearly needs to be reviewed to make it more manageable and cost efficient. Suggestions that could be explored include:

- setting common submission dates across all subjects within a school
- submission of all assessment materials to NZQA instead of individual moderators
- adopting a lighter sampling approach

- reducing the administrative demands by reducing and reviewing the moderation pro-formas
- seconding full time moderators
- increased provision of pre-moderated assessment materials.

8.3 What are the likely consequences of national implementation for students?

The present study investigated the impact of assessment against the Level 2 PUS on students' workload, unit standards-based reporting and motivation. This section draws on the findings of the present study and the research literature to discuss what the likely consequences will be for students if they are assessed against US for all their subjects in years 11-13.

The present study found that students who were assessed against the Level 2 PUS experienced a higher frequency and larger number of smaller assessments that carried less weighting compared to SFC assessments. This finding is not specific to Level 2 physics but is a systemic consequence of assessment close to learning and the large number of US, elements and performance criteria for each school subject. If students were to be assessed against US for all their school subjects, the cumulative effect would be a large number and frequency of assessment and an increase in the total class time devoted to assessment. This has the potential to adversely affect students' learning because they are constantly concentrating on meeting short-term goals, possibly at the expense of longer-term goals such as synthesis and integration. In addition the demands of continuous assessment and coordination of assessment dates may become too cumbersome and stressful for students. The NZPPTA Inquiry report (NZPPTA, 1997a) concurs with this conclusion and argues that the US system cannot succeed if it involves unacceptable levels of workload and stress for students.

The system of assessment against US was initially developed to allow reporting of achievement that enables subjects to stand-alone and grades to be awarded without reference to student performance in other subjects or levels. The present study found that the system for reporting achievement on assessments against the Level 2 PUS was transparent and that students liked receiving credit at the unit level, the absence of scaling, the independence of subjects and the seamlessnes of the NQF. The diagnostic nature of assessment against US coupled with the opportunity for reassessment was found to encourage thorough learning in the assessed domain. These aspects relate to the structure of the assessment system and are not subject specific to physics and are therefore likely to apply across other subjects and levels to students in general.

The study found however that there were problems associated with the competency-based nature of the reporting system. Aspects that students did not like included the all or nothing system of reporting achievement and the lack of recognition of different levels of achievement. The pass/fail nature of the competency-based approach to grading student achievement in assessments against US was unpopular with teachers and students. It adversely affected the motivation of students whose work clearly surpasses the standard and did not enable the recognition of excellence. The systemic failure of the competency-based approach to recognise excellence has also been commented on by Austin (1997), NZIP (1996), NZPPTA (1997a), Rawson (1997), Salter and Hayden (1996) and Sinclair (1997).

Conversely the study found that the level of achievement required for credit of some US was set too high and made it difficult for less able students to gain recognition for their learning. This adversely affected the motivation of students who did not attain success and found it too difficult to attain credit on the framework. This finding applies more widely to other subject areas. The NZPPTA Inquiry (NZPPTA, 1997a) and Vlaardingerbroek (1996) concluded that the level for credit is set too high for some US.

The NZPPTA Inquiry (NZPPTA, 1997a: 54) reached the conclusion:

> Perhaps the most important point to make here is that the system fails to recognise that while good students can be fairly successful in their understanding, they are rarely perfect and that weaker students, though never perfect, still succeed to some extent.

Extending the pass/fail nature of reporting to include different levels of achievement may resolve this issue. A possible approach is to have three levels for reporting student achievement ranging from pass to merit and excellence. The level of achievement required for competency should be set at a level that is achievable by the majority of students. To enable this system to be adopted the US would need to be redesigned to include separate performance criteria for each of these achievement levels. These criteria would need to be scaffolded so that the criteria for each level of achievement incorporate the requirements of the criteria for lower levels. This approach incorporates some aspects of ABA discussed in Chapter 1 (§1.2.3). This type of approach is suitable for wider implementation across other subjects and is supported by the NZPTA Inquiry (NZPPTA, 1997a: 103) which concluded:

> ... unit standards would be better suited to a model offering a range of grades than to the binary pass/fail model. We favour making available up to two levels beyond pass and suggest that those levels be labeled 'pass with merit' and 'pass with excellence'.

Any changes to the grading system would need to be illustrated by exemplars of student work and supported by training to help teachers make accurate and consistent judgements regarding the levels of student achievement.

The overall conclusion about the likely consequences of national implementation of assessment against US is that while the system has advantages for students there are a number of concerns that need to be addressed before the system is fully implemented. The workload for students that results from continuous internal assessment may become unmanageable and place undue stress on students. Introducing a mixture of internal and external assessment and reducing the number of US in each subject could alleviate this. While the competency-based nature of reporting may be suitable for students of average ability it adversely affects the motivation and learning of lesser and more able students. A broadening of the system of to include recognition of different levels of achievement may address this concern.

8.4 What are the likely consequences of national implementation for teachers?

This section discusses the likely consequences for teachers if assessment against US is fully implemented for all subjects at all levels of secondary schooling. These consequences relate to:

- teachers' workload
- the impact on classroom practice
- the quality of professional development and resources

a) Teachers' workload

The workload associated with the implementation of assessment against US is a major concern for teachers Coutts and Mc Alpine, 1996; Rosser 1996; Salter and Hayden, 1996). For many teachers the full-scale implementation of assessment against US in all subjects across levels 1-3 of the Framework necessitates assessment against US at more than one level and possibly in more than one curriculum area. This will seriously impact on teachers' workload. The present study identified that the extra workload was due to the design of new assessment tasks and schedules, reassessment, the administrative demands of record keeping liaison with possibly more than one moderator, attendance at local provider meetings and training in standards-based assessment.

In the final year of the study the workload associated with assessment against the level 2 PUS had settled and was similar to that associated with assessment for SFC. This finding is similar to Meldrum's (1995) finding that the workload is initially considerable but becomes more manageable over time as teachers develop familiarity with the US and assessment and moderation procedures. Since both SFC and US are completely internally assessed, the replacement of SFC by US may not significantly add to teachers' workload over the longer term.

The impact on workload will be particularly felt in Years 11 and 13 where the external summative SC and Bursary examinations will be replaced by internally assessed level 1 and 3 US. The introduction of a mixture of internally and externally assessed US would ameliorate this. The decision whether an individual unit standard should be internally or externally assessed should be based on validity considerations. The practical and performance components of subjects are more validly assessed internally whereas the theoretical components can be more reliably assessed externally.

Participation in a moderation system is very time consuming and involves moderators and teachers in a large amount of administration. There is a large number of moderation pro-formas to be completed and there is some duplication between forms. These forms need to be reduced in number and redesigned to make them less cumbersome and more user friendly for teachers.

The management of assessment of individual performance in a whole class setting has manageability implications. Since authenticity and sufficiency of evidence must be assured there is pressure on equipment and time for students to be individually assessed during class time.

Given the widespread debate and industrial action it appears doubtful that teachers are willing or enthusiastic to sustain the additional workload associated with full implementation, particularly at a time when other developments such as the introduction of new curricula and attestation systems are placing additional demands on teachers. This conclusion concurs with the NZPPTA (NZPPTA, 1997a: 111) Inquiry's conclusion that:

> ... operating the Framework style assessment model as it currently exists and at all three levels (Years 11, 12 and 13) will probably produce an unmanageable workload in the transitional period and at least in the medium term.

b) the impact on classroom practice

The present study shows physics' teachers' opinion on the impact of assessment against the level 2 PUS on teacher enthusiasm and classroom practice was divided. Teachers were concerned that the highly prescriptive nature of the PUS and the continuous assessment process tended to drive classroom practice at the expense of the curriculum. Positive impacts on classroom practice included clearly specified learning outcomes, the diagnostic and formative nature of Framework assessment and the professional development associated with the implementation of assessment against US. The seamless nature of the Framework allowed for flexibility in course design and the validation of school-based courses that were not recognised under the existing exam system. The sample for the current research consisted of committed teachers that assessed against US voluntarily. More adverse views could be expected to prevail if all teachers were involved in the system. The implication of this is that careful consideration needs to be given to managing the factors that teachers identified as impacting positively and negatively on teacher enthusiasm and classroom practice.

c) The quality of resources and professional development

Coutts and McAlpine (1996) argue that long-term adequate resourcing is essential to maintain a high level of quality assurance and public confidence in the Framework. The resources that are most relevant for teachers include assessment resources and professional development.

Access to quality resources helps boost teacher confidence in implementing the system, exemplify the standards and reduce workload. Teachers value assessment resources such as the assessment guides, pre-moderated activities and CDs of assessment activities. The provision of comprehensive resources such as item banks of pre-moderated assessment activities accompanied by exemplars of student work that clearly illustrate the standard is essential in the implementation period. Since the credibility of the new system is linked to the quality and accuracy of the resources there should be thorough checks before these are released for general use. The level of checking, moderation and presentation should be commensurate with that for external examinations.

The training and funded cluster meetings organised by the NZQA were viewed positively by teachers. There is a need for ongoing training and cluster meetings in the initial implementation period. Mechanisms for the provision and funding of this training need to be established.

8.5 What are the likely consequences of national implementation for schools?

The full implementation of assessment against US for all subjects at all levels of the senior secondary school has implications for school organisation, quality management systems and professional development budgets. In addition there is a need for a definite and realistic implementation timeline to enable schools to engage in long-term planning.

Prior to the introduction of the Framework there were many school-based Level 2 and 3 courses that did not receive national recognition. Typical courses included electronics, tourism and hospitality and catering. These courses are now recognised on the Framework. In addition, the flexibility of the Framework and the modular nature of US enable schools to design school-based courses by combining US from a variety of subject areas. While this caters for students' educational needs it creates timetabling pressures and has an impact on enrollments in traditional courses. Additional pressures on the timetable arise from students passing some US courses at level 1 or 2 and then expecting to enrol for some US at a higher level. Vice versa if students do not complete all of the US that make up a course there is pressure for them to repeat these US the following year or for schools to practice multi-level assessments in the same year group. Under the exam system students who do not pass a course are expected to repeat the whole years' work. These pressures combine to challenge the existing course and timetable structures and may force schools into a modular rather than year long course timetable structure.

Since high stakes assessment will be placed at the school rather than national level it imposes additional demands on school quality assurance systems. This requires more detailed assessment policies and increased communication with NZQA. Additional assessment coordinator and NZQA liaison positions may need to be created and ancillary staffing levels increased to deal with the administrative demands associated with the increased level of internal assessment. The effects of this will be particularly felt at Year 11 and 13 where internal assessment will replace external examinations.

The assessment of performance skills for individual students such as practical and project work will place additional demands on resources and teacher workload. In particular it poses challenges for ensuring that work carried out in groups or extended investigations and research projects that are completed outside school hours are entirely the student's own work. Coupled with this is the issue of sufficiency of evidence. This relates to how many times students have to demonstrate competence before credit is awarded.

These issues need to be addressed in school assessment policies and some national guidelines need to be formulated for schools. Audits of school assessment policies could be used to determine whether schools have adequate procedures in place to handle the additional demands of standards-based high stakes internal assessment for national qualifications at all levels of the senior secondary school.

Teachers who are entering or returning to the profession and teachers from overseas need to be trained in standards-based assessment. Ongoing training opportunities will need to be provided for those teachers. This is not provided by NZQA and will need to be sourced and paid for by schools.

The political and industrial uncertainty associated with the implementation of assessment against US has made long-term planning for schools difficult and necessitated the need for dual assessment to satisfy the demands of both standards-based and norm-referenced assessment. Since dual assessment often involves dual marking of assessment activities or running parallel assessment systems, it is unpopular with both teachers and students and adds considerably to the workload and stress for teachers and students. The government needs to set a definite timeline for transition to the new standards-based system to avoid prolonged periods of dual assessment. If the government does not set a definite timeline or the proposed implementation dates keep changing it is possible for a culture of cynicism and lack of commitment to change to develop in the teaching community. The NZPPTA Inquiry (NZPPTA, 1997a: 113) concluded:

> A properly planned and funded transition would allow measured and thorough establishment of the new system over the course of three or four years.

8.6 Conclusion

The discussion of the implications of the current research for the implementation of the NQF in secondary schools shows that there are definite advantages associated with the paradigm shift towards standards-based assessment. The NZPPTA Inquiry (NZPPTA, 1997a: 101) concluded:

> The Inquiry has accepted that standards-based assessment is more desirable on educational grounds than norm-based assessment.

This does not mean that the present system should be adopted without review. The discussion highlighted a number of validity, reliability and manageability concerns related to the present system of competency-based assessment against US. These issues need to be addressed in a comprehensive review of the system before it is nationally implemented across all levels and subjects in the senior secondary school. Chief recommendations for improving the system include a reduction in the number of US in each subject area, fewer and broader performance criteria and the inclusion of scaffolded criteria to measure different levels of achievement including excellence. The introduction of a mixture of internally and externally assessed US would make the system more manageable and contribute towards achieving comparability between schools. While effective, the quality assurance system needs to be streamlined to make it more manageable and affordable. Lastly a definite but realistic timeline needs to be established for the full implementation of the Framework in the senior secondary school.

Any future changes to the national assessment system for levels 1-3 of the Framework would need to be fully evaluated in the context of each of the essential learning areas of the New Zealand Curriculum Framework. The system for the evaluation of the quality of assessment established by the present study could be more widely adapted for this purpose.

References

Abacus Concepts. (1994). *Statview*. Berkely: Abacus Concepts.

Austin, L. (1996). Value of Standards far from Proven. *New Zealand Education Review*, August 8-14.

Austin, M. (1995, May 30). Report misleads. *The Press*.

Barker, A. (1996, March 30) in *New System 'Aids Capable Students'. The Press.*

Batchelor, G. (1996, November 27). Physics Students Tell why they do not Like the New Assessment System. *The Press*.

Bennetts, J., Fairbrother, R.W. and Willmott, A.S. (1986). The Moderation of Teacher Assessment. *Education in Science*, No 120, Nov. pp.16-18.

Bennetts, J., Fairbrother, R.W. and Willmott, A.S. (1987). The Moderation of Teacher Assessment: Moderation by Inspection. *Education in Science*, No 122, April. pp. 11-15.

Best, J. W. and Kahn, J. V. (1998). *Research in Education*. Massachusetts: Allyn and Bacon.

Board of Studies. (1988). *Newsletter No 1*. Wellington: Board of Studies. August.

Braun, L. (1996). in Framework Issues Highlighted. *New Zealand Education Review*, August 30.

Broadfoot, P. (1990) in Peddie, R. A. and Tuck, B. (1995). *Setting the Standards: Issues in Assessment for National Qualifications*. Palmerston North: Dunmore Press Ltd.

Brock-Utne, B. (1980). What is educational research? *Classroom Action Research Network Bulletin*, no 4, pp. 10-15.

Burns, R.B. (1994). *Introduction to Research Methods*. Melbourne: Longman Australia.

Burke, J. and Jessup, G. (1990). Assessment in NVQ's: disentangling reliability from validity. in Horton, T. (ed.). *Assessment debates*. London: Hodder and Stoughton.

Carter, M. (1996). *An Open Letter on the Subject of Unit Standards.* Auckland: Diocesan school for Girls.

Chamberlain, J. (1996). Has the Qualifications Authority Duped the Nation? *North and South,* September.

Chapman, R. (1996) in Hotere, A. Teachers now Dubious of Unit standards. *New Zealand Education Review*, 4 October.

Chapple, G. (1996). Elite Schools Waylay a Revolution. *The Listener*, July 13.

Childs, R. (1995). *Unit Standards- Barking up the Wrong Assessment Tree?* New Zealand History Teachers' Association, April.

Clarke, S. (1993). In Search of the National Standard. *British Journal of Curriculum and Assessment*. (Vol. 3, No 3).

Codd, J., McAlpine, D. and Hansen, S. (1990). *Achievement, Assessment and Consensus: An evaluation of the Sixth Form Certificate Practical Art Moderation Trial.* Palmerston North: Massey University, Department of Education.

Codd, J., McAlpine, D. and Poskitt, J. (1995). *Assessment Policies in New Zealand.* Palmerston North: The Dunmore Press.

Coogan, P. (1996). *Standards-based Assessment: Lessons for New Zealand from the United Kingdom*. London: University of London, Institute of Education.

Coutts, C. and McAlpine, D. (1996). *Monitoring the Qualifications Framework Trials in Schools - Evaluation of the External Moderation System*. Palmerston North: Educational Research and Development Centre, Massey University.

Cowling, L. (1994). Issues in Moderation of National Curriculum Assessment at Key Stage 3. *Education in Science*, no 158, June.

Crooks, T. (1990). *Evaluation of Sixth Form Certificate Training in Assessment Developments*. University of Otago Higher Education Development Centre, June 1990.

Crooks, T., Kane, M. and Cohen, A. (1996*). Threats to the Valid use of Asessments*. Unpublished manuscript.

Davidson, C. and Steer, M. (1989). *Report of an investigation into Teacher Panel Moderation and the Use of Grade-related Criteria associated with Sixth Form Certificate Geography*. Wellington: Department of Education.

Department of Education. (1986). *Learning and Achieving: Second Report of the Committee of Inquiry into Curriculum Assessment and Qualifications in Forms 5 to 7.* Wellington: Department of Education.

Department of Education. (1989). *Visitation as a Moderation Method.* Wellington: Department of Education, Quality and Assessment Division.

Elley, W.B. and Livingstone, I. D. (1972). *External Assessments and Internal Examinations.* Wellington: New Zealand Council for Educational research.

Elley, W. B. (1988). *Moderation by Means of Reference Tests.* Christchurch: University of Canterbury, May.

Elley, W.B. (1994, November 17). Fundamental Flaw at the Heart of the New Qualifications Plan. *The Press.*

Elley, W.B. (1996, June 21). Enough of Fudge and Evasion, the NZQA's Plans just will not do. *The Press.*

Elsdon, K., T. (1987) External Examining and Moderating: A Tyro's Reflections. *National Association for Staff Development Journal*, Nov-Dec.

Frederiksen, J. and Collins, A. (1989). A system's approach to educational testing. *Educational Researcher*, 18, 9, pp. 27-32.

Gay, L.R. (1996). *Educational Research.* Upper Saddle River, N.J.: Merrill.

Gernhoefer, A. (1996). in Chamberlain, J. Has the Qualifications Authority Duped the Nation? *North and South*, September.

Gilmore, A. (1991). *An Evaluation of the Sixth Form Certificate English Moderation Trial.* Christchurch: University of Canterbury.

Gipps, C. (1994). *Beyond Testing: Towards a Theory of Educational Assessment.* London: The Falmer Press

Gipps, C. and Stobart G. (1994). *Assessment: a Teachers' Guide to the Issues.* Hodder and Stoughton.

Gipps, C. (1990). *Assessment. A Teacher's Guide to the Issues.* London: Hodder and Stoughton.

Griffin, P.and Nix, P. (1991). *Educational Assessment and reporting: A New Approach.* Sydney: Harcourt, Brace and Jovanovich.

Gronlund, N., E. (1973). *Preparing Criterion Referenced Tests for Classroom Instruction.* New York: Macmillan.

Hansen, J. (1997). *Milestone not Millstone! 1996 history trials schools milestone five survey.* Wellington: NZQA.

Harlen, W. (ed). (1994). *Enhancing Quality in Assessment.* London: Paul Chapman Publishing Ltd.

Hawke, G.R. (1988). *Report on the Working Group on Post-compulsory Education and Training.* Wellington: Office of the Associate Minister of Education.

Hearn, S. (1996a). Overseas Lessons: Standards-based Assessment in England and New Zealand. *PPTA News*, December.

Hearn, S. (1996b). *Standards-based assessment in New Zealand and the United Kingdom: The New Zealand qualifications framework and GNVQ.* University of London, Institute of Education.

Hearn, S. (1997a) in Hotere, A. Education Qualifications Policy due in April. *New Zealand Education Review*, February 19.

Hearn, S. (1997b). in Secondary Teachers Target Framework Reforms. *Eduvac: The Education Weekly*, February 24.

Hood, D. (1996, August 27). in Schools Using new Framework. *The Press.*

Hughes, D. and Lauder, H. (1990). Public examinations and the structuring of inequality: A case study. In Lauder, H. and Wylie, C. (eds.). *Towards Successful Schooling.* London: The Falmer Press, pp 151-173.

Irwin, M. (1994). *Curriculum, Assessment and Qualifications: An Evaluation of Current Reforms.* Auckland: Education Forum, May.

Irwin, M. (1995, June 1). The Mistake of Trying to Force Education into one Framework. *The Press.*

Irwin, M., Elley, W., and Hall, C. (1995). *Unit Standards in the National Qualifications Framework.* Auckland: Education Forum, May.

Izard, J. (1991). *Assessment of Learning in the Classroom*, Geelong: Deakin University.

James, M. and Conner, C. (1993). Are Reliability and Validity Achievable in National Curriculum Assessment? Some observations on Moderation at Key Stage 1 in 1992. *The Curriculum Journal.* Vol 4, no 1.

Kempa, R. (1986). *Assessment in Science.* Cambridge : Cambridge University Press.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Leach, M. (1996, October 29) in, Teachers' Decision Upsets Qualifications Authority. *The Press*.

Lee, G. D and Lee, H. F. (1992). *Examinations and the New Zealand School Curriculum: Past and Present*. Delta Research Monograph No 12. Palmerston North: Massey University, Education Department.

Linn, R. L. and Gronlund, N.E. (1995). *Measurement and Evaluation in Teaching*. New York: Englewood Cliffs.

Mann, B. (1997). *Discussion Paper on the NZQA and its effect on General Education*, Christchurch: St Margaret's College.

Metcalfe, J. (1996). Editorial: The unit standards trials-tick or cross? Unpublished paper submitted for publication in *Chemistry New Zea*land.

Methven, P., Goddard, P. and Thompson, T. (1996). *The National Qualifications Framework: Issues*. Unpublished paper prepared for NZQA, Wellington.

Mc Cormick, R. and James, M. (1983). *Curriculum Evaluation in Schools*. London: Routledge.

McIntyre, R. (1995). *Moderation of Draft Mathematics Unit Standards*. End of year report to NZQA moderation services, Wellington.

McKay, E. and Peters, M. (1989) *An Evaluation of the Sixth Form Certificate Home Economics Moderation Trial*. Auckland College of Education and University of Auckland.

Marshall, G. (1994, September 28). Optimism Lack for NZQA Plan. *The Evening Post*.

Matheson, D. (1997). NZQA advice discarded by schools. *New Zealand Education Review*, April 16, P5.

Masters, G. N. and Mcbryde, B. (1994) *An Investigation of the Comparability of Teachers' Assessment of Student Folios*. Research Report Number 6. Brisbane:Tertiary Entrance Procedures Authority, April.

Messick, S. (1989) in Gipps, C. (1994). *Beyond Testing: Towards a Theory of Educational Assessment*. London: The Falmer Press

Ministry of Education (1990). *Assessment for Better Learning*. Wellington: Learning Media.

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

Ministry of Education. (1993b). *Science in the New Zealand Curriculum*. Wellington: Learning Media.

Ministry of Education. (1994a). *Physics in the New Zealand Curriculum*. Wellington: Learning Media.

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

Ministry of Education. (1995). *Directory of New Zealand Schools and Tertiary Institutions*. Wellington: Ministry of Education.

Ministry of Education. (1997). *A Future Qualifications Policy for New Zealand: A Plan for the National Qualifications Framework*. Wellington: Ministry of Education , June 5.

Ministry of Education. (1998). *Achievement 2001, Qualifications for 16 to 19 year olds*. Wellington: Ministry of Education.

Morris, J. in Hotere, A. (1996). Schools seek Framework Review. *New Zealand Education Review*, May 8.

New Zealand Council for Educational Research. (1987). *The Best of Set: Assessment*. Wellington: New Zealand Council for Educational Research.

New Zealand Government. (1988). *Tomorrow's Schools: The Reform of Education Administration in New Zealand*. Wellington: Government Printer.

New Zealand Government. (1989a). *Learning for Life: Education and Training Beyond the Age of Fifteen*. Wellington: Government Printer.

New Zealand Government. (1989b). *Learning for Life Two: Education and Training Beyond the Age of Fifteen- Policy Decisions*. Wellington: Government Printer.

New Zealand Institute of Physics. (1996). *Education Sub-committee Survey on Unit Standards: Summary of Results*. Christchurch: NZIP Education sub-committee.

New Zealand Post Primary Teachers' Association. (1967). *PPTA Journal*. Vol 14, No 3 p. 25. Wellington: New Zealand Post Primary Teachers' Association.

New Zealand Post Primary Teachers' Association. (1971). *The Journal*. Vol. 18 No 7 p. 30. Wellington: New Zealand Post Primary Teachers' Association.

New Zealand Post Primary Teachers' Association. (1992). Principals Response to MSG Announcement. *PPTA News*, Vol. 13, No 7.

New Zealand Post Primary Teachers' Association. (1993). Editorial. *PPTA News*, Vol. 14, No 1.

New Zealand Post Primary Teachers' Association. (1996). A Warrant of Fitness for the Frameworks. *PPTA News*, November.

New Zealand Post Primary Teachers' Association. (1997a). *Te Hiro Hou: Report of the Qualifications Framework Inquiry.* Wellington: New Zealand Post Primary Teachers' Association.

New Zealand Post Primary Teachers' Association. (1997b). Jury still out on the Framework. *PPTA News* vol. 18, no 8.
New Zealand Qualifications Authority. (1990). *Towards a National Qualifications Framework.* Wellington: New Zealand Qualifications Authority.

New Zealand Qualifications Authority. (1991a). *Achievement Based Assessment in New Zealand.* Wellington: New Zealand Qualifications Authority.

New Zealand Qualifications Authority. (1991b). *Designing the Framework.* Wellington: New Zealand Qualifications Authority.

New Zealand Qualifications Authority. (1992). *Designing a Moderation System.* Wellington: New Zealand Qualifications Authority.

New Zealand Qualifications Authority. (1993). *Quality management Systems for the National Qualifications Framework.* Wellington: New Zealand Qualifications Authority.

New Zealand Qualifications Authority. (1994). From Bouquet to Brickbats. *Learn: The Magazine for Education and Training Professionals.* August.

New Zealand Qualifications Authority. (1995a). *Physics Unit Standards.* Wellington: NZQA.

New Zealand Qualifications Authority. (1995b). *Standards-based Assessment for the national Qualifications Framework: Science Facilitator's Handbook,* Wellington: NZQA.

New Zealand Qualifications Authority. (1995c).*Role definition: Local Moderator.* Wellington: NZQA. (issued to local moderators)

New Zealand Qualifications Authority. (1995d). *Role definition: Regional Moderator.* Wellington: NZQA. (issued to regional moderators)

New Zealand Qualifications Authority. (1995e). *Role definition: National Moderator.* Wellington: NZQA. (issued to national moderators)

New Zealand Qualifications Authority. (1996a). Why was the Qualifications Framework Created? Trial Newsletter 10. Wellington: NZQA, November.

New Zealand Qualifications Authority. (1996b). *Standard-Based Assessment for the National Qualifications Framework: Physics.* Wellington: NZQA.

New Zealand Qualifications Authority. (1996c). *Assessment Guide: Physics.* Wellington: NZQA.

New Zealand Qualifications Authority. (1997a). *Report on the 1996 School Trials Surveys.* Wellington: NZQA.

New Zealand Qualifications Authority. (1997b). *Standard-Based Assessment for the National Qualifications Framework: Physics* (second edition). Wellington: NZQA.

New Zealand Qualifications Authority. (1998). *School Qualifications Handbook: Regulations, Prescriptions.* Wellington: NZQA.

Neyland, J. (1994). *Problems with the Proposed Unit Standard Framework for Mathematics*, Wellington College of Education, 22 May.

OCEA (1985). *Pilot Phase Materials.* University of Oxford Delegacy of Local Examinations.

Peddie, R. A. (1990). *Assessing Alone: An Evaluation of the Sixth Form Certificate French Trial, 1989.* Auckland: University of Auckland Education Department.

Peddie, R. A. (1992). *Beyond the Norm? An introduction to Standards-Based Assessment.* Wellington: New Zealand Qualifications Authority.

Peddie, R. A. and Tuck, B. (1995). *Setting the Standards: Issues in Assessment for National Qualifications.* Palmerston North: Dunmore Press Ltd.

Pepper, J. (1995). *Moderation of Draft Geography Unit Standards.* End of year report to NZQA moderation services, Wellington.

Picot, B. (chair). (1988). *Administering for Excellence: Effective Administration in Education.* Wellington: The Task force to Review Education Administration.

Radnor, H. and Shaw, H. (1995). *Developing a Collaborative Approach to Moderation.* in Torrance, H. *Evaluating Authentic Assessment: Problems and Possibilities in new Approaches to Assessment.* Buckingham: Open University Press.

Rawson, G. (1997). Personal communication. NZQA, 18 June 1997.

Rosser, M. (1996, December 3). College Head Slams Assessment System. *Christchurch Star.*

Sadler, D. R. (1986). *School-based Assessment and Autonomy.* Discussion paper 11. Brisbane: Board of Secondary School Studies.

Sadler, D. R. (1993). *An Exploration of Aspects of Comparability in the Assessment of the Achievement by Students in Queensland Secondary Schools.* Brisbane: Tertiary Entrance Procedures Authority.

Sadler, K. (1994). A suggested Model for the audit Moderation of Teacher Assessment at Key Stage 2. *British Journal of Curriculum and Assessment.* Vol. 4, No3. pp31-32,36.

Salter, D and Hayden, D. (1996). Not Very Great Expectations. *New Zealand Education Review*, November 1.

Sass, R. E. and Wagner, G. A. (1992). *Moderation Strategies for Achieving Consistency with Unit Standards.* Wellington: NZCER.

Secondary Examinations Council (1986). *Working Paper 3: Policy and Practice in School-based Assessment.* London: SEC.

Sewell, K. in Hotere, A. (1996). Schools seek Framework Review. *New Zealand Education Review*, May 8.

Shorrocks, D., Daniel, S., Frobisher, L., Nelson, N., Waterson, A., and Bell, J. (1992). ENCA. *Project Report.* London: SEAC.

Sinclair, S. (1997) Use the Force Luke. *Burnside High School Terminal*, June.

Smithers, A. (1997). *The New Zealand Qualifications Framework.* Auckland: Education Forum.

Snook, I. and St George, R. (1986). Failing School Certificate. *Delta*, 37, pp. 25-31.

Squire, S. (1996). *Physics Unit Standards: a personal evaluation.* Unpublished paper.

St George, R. (1985). Describing School Certificate: Is there something fishy about this scaling? *PPTA Journal*, Term 3, PP. 40-42 and 44-47.

Sutton, R. (1994). Roadworks Ahead. *Learn: The Magazine for Education and Training Professionals.* Number 3, December.

Tarling, N. in Matheson, D. (1996). Form Seven Hybrid. *New Zealand Education Review*, November 29.

Thomas, W. (chair). (1944). *The post-primary school curriculum: report.* Wellington: Consultative Committee on the Post –Primary School Curriculum.

Thorndike, R.M., Cunningham, G.K., Thorndike, R.L. and Hagen, E.P. (1991). *Measurement and Evaluation in Psychology and Education.* New York: Mac Millan Publishing Company.

Torrance, H. (1996). in Assessment Doesn't Mean Improvement. *New Zealand Education Review*, June 19.

Torrie, I. (1995). *Moderation of Draft Chemistry Unit Standards.* End of year report to NZQA moderation services, Wellington.

Torrie, I. (1997). *Chemistry Moderation Newsletter # 2.* Ian Torrie 23 March 1997.

Viviani, N. (1990). *The Review of Tertiary Entrance in Queensland.* Brisbane: Department of Education.

Vlaardingerbroek, B. (1996). The poor Relation. *New Zealand Education Review*, November 1.

Walker, L. (1990). *The Assessment of Practical Work for SFC Biology, Chemistry and Physics in Christchurch Schools in 1988*. Christchurch: University of Canterbury Education Department. Research Report No 90-3.

Willmott, A.S., Bennetts, J. and Fairbrother, R.W. (1987). The Moderation of Teacher Assessment: Consensus Moderation. *Education in Science*, No 121, Jan.

Wilson, B. (1996). Equal and Oppposite Reactions. *New Zealand Education Review*, November 8.

Woods, M. (1992). Assessment in New Zealand Secondary Schools, in *Qualifications for the 21st Century*. Wellington: New Zealand Qualifications Authority.

Appendices

Appendix 1: Achievement Based Assessment Criteria for Sixth Form Certificate Physics

	Level 1	Level 2	Level 3	Level 4	Level 5
Personal Skills	Attempts some tasks with constant supervision.	Completes most tasks with some supervision	Completes tasks to a high standard with supervision.	Accepts responsibility for the completion of tasks without supervision.	Independently completes tasks to a high standard, and shows leadership skills in a group situation.
Physical Ideas	Recalls some physical ideas.	Recalls accurately a wide range of physical ideas.	Relates physical ideas to familiar situations.	Demonstrates an understanding of physical ideas by using them to explain familiar situations.	Demonstrates a full understanding of physical ideas by using them to explain familiar and unfamiliar situations.
Planning Investigations	Identifies variables to be investigated.	Presents procedures appropriate to the investigation.	Presents a plan for the investigation.	Presents a coherent plan that could lead to valid conclusions.	Presents a comprehensive and coherent plan that would lead to valid conclusions, and attempts to anticipate potential problems.
Gathering Information/Data	Describes the purpose of gathering information/data.	Gathers specified information/data from given resources or equipment.	Selects appropriate resources and equipment and gathers specified information/data.	Independently gathers appropriate information/data and presents it clearly.	Independently gathers all relevant information/data, presents it in precise form and evaluates its limitations.
Interpretation	Identifies some trends in information/data.	Attempts to analyse trends in information/data.	Analyses information/data and attempts to draw conclusions.	Analyses information/data and draws valid conclusions.	Draws valid conclusions from information/data and demonstrates an understanding of their limitations.
Communication	Attempts to express ideas.	Expresses ideas using a range of forms.	Expresses ideas clearly using a range of forms.	Expresses ideas clearly using an appropriate form.	Expresses ideas clearly and concisely using the most appropriate forms.

To use these broad descriptors in practice they have to be translated into specific levels of achievement that apply to the context of particular assessment activities. For instance, the following marking schedule is based on the grade related criteria for assessing student understanding of physical ideas. It was used as part of a sixth form physics assessment programme to assess the various levels of student performance for a physics project.

Marking schedule for assessing student understanding of physical ideas for a sixth form physics project

Level 1:	*Very limited use of physics concepts or knowledge in the project.*
Level 2:	*The project contains some physical concepts but there are weaknesses in the student's use and understanding of these.*
Level 3:	*The student has identified physical ideas and linked them appropriately with the situation in this project.*
Level 4:	*The student has identified physical ideas and uses these correctly explain the observations in the project.*
Level 5:	*The student has identified physical ideas, and correctly uses these to explain the observations in the project and a range of additional situations.*

Appendix 2

The 1997 Physics Unit Standards title matrix

The table shows the 1997 title matrix and credit values for the level 1-3 Physics Unit Standards. This includes all the Physics Unit Standards that may be assessed against in the senior secondary school.

Outcomes derived from Level 6 of *Physics in the New Zealand Curriculum* and Level 6 of *Science in the New Zealand Curriculum* ↓ NQF LEVEL ONE YEAR 11	Outcomes derived from Level 7 of *Physics in the New Zealand Curriculum* ↓ NQF LEVEL TWO YEAR 12	Outcomes derived from Level 8 of *Physics in the New Zealand Curriculum* ↓ NQF LEVEL ONE YEAR 13
6366: Physics Demonstrate knowledge of linear motion. (2)*	6378: Physics Demonstrate knowledge of motion in one and two dimensions. (4)	6397: Physics Demonstrate an understanding of circular rotational and simple harmonic motion. (6)
6367: Physics Explain the results of forces acting on an object (2)	6379: Physics Demonstrate understanding of energy, momentum and equilibrium. (4)	6388: Physics Apply formulae, graphical, vectorial and phasor methods to find unknowns for a physical system. (3)
6368: Physics Demonstrate knowledge of energy transformation (2)	6380: Physics Apply formulae, graphical and vectorial methods to find unknowns for a physical system. (3)	6389: Physics Describe and determine unknowns for direct current electrical systems. (3)
6369: Physics Apply formulae and graphical methods to find unknowns for a physical system (3)	8770: Physics Describe, construct and determine unknowns for electromagnetic systems (4)	6390: Physics Describe and determine unknowns for alternating current electrical systems. (4)
6370: Physics Describe and construct simple electrical systems. (3)	8769: Physics Describe, construct and determine unknowns for electrical systems (4)	6391: Physics Demonstrate an understanding and determine unknowns for wave systems. (4)

6371: Physics Describe and determine unknowns for heat and light systems. (4)	6382: Physics Demonstrate knowledge of waves. (4)	6392: Physics Analyse the development of a selected area of physics and a physics-based application. (2)
6372: Physics Investigate and describe the influences physics based applications have on peoples lives. (2)	6383: Physics Describe the development of a selected physics idea and a physics based application. (2)	6393: Physics Investigate a physical system to determine a relationship with guidance. (4)
6373: Physics Investigate a physical system to determine a relationship with direction. (4)	6384: Physics Investigate a physical system to determine a relationship with supervision. (4)	6394: Physics Carry out a practical investigation of a physics-based application with guidance. (3)
6374: Physics Carry out a practical investigation of a physics based application with direction. (3)	6385: Physics Carry out a practical investigation of a physics based application with guidance. (3)	6395: Physics Use graphical analysis to determine non linear physical relationships. (4)
6375: Physics Use graphical analysis to recognise a directly proportional physical relationship. (2)	6386: Physics Use graphical analysis to determine simple non linear physical relationships. (3)	6396: Physics Describe and discuss models of atomic systems. (3)
6376: Physics Demonstrate knowledge of atomic structure and fission reactions. (2)	6387: Physics Demonstrate knowledge of elementary nuclear physics and radioactivity. (2)	
6377: Physics Observe and explain the movement of objects in the solar system (3)		
8767: Physics Demonstrate knowledge of heat and temperature. (4)		

* Number of credits

Appendix 3

Elements 1 and 2 of Physics Unit Standard 6380

Unit No: 6380
Issued: 01 Nov. 96

PHYSICS

Apply formulae, graphical and vectorial methods to find unknowns for a physical system

level:	2
credit:	3
final date for comment:	October 1997
expiry date:	December 1998
sub-field:	Science
purpose:	People credited with this Unit Standard are able to: recognise the principles and quantities of a physical system; determine an unknown quantity using a formula; determine an unknown quantity using a graph; and determine an unknown quantity using two dimensional vectors.
entry information:	Open
accreditation option:	Evaluation of documentation by NZQA
moderation option:	A centrally established and directed external moderation system has been set up by NZQA on behalf of the Science and Technology Advisory Group
special notes:	1. Situations involving the finding of an unknown quantity for the assessment of the outcomes of this Unit Standard must be consistent with the achievement objectives, possible learning experiences, content and assessment examples from level 7 of: Ministry of Education, *Physics in the New Zealand Curriculum* (Wellington: Learning Media, 1994). 2. The term 'system' is used to cover phenomena and devices.

Elements and Performance Criteria

element 1
Recognise the principles and quantities of a physical system..
performance criteria
1.1 Principle/s identified for the system will enable an unknown to be found.
1.2 The values of the physical quantities are identified with their symbols and Systeme Internationale (SI) units.

element 2
Determine an unknown quantity using a formula.
performance criteria
2.1 A formula is selected which is appropriate to the quantity to be determined and valid working is shown.
2.2 The determined quantity is consistent with the information supplied.

2.3 The answer is stated using a complete sentence, an appropriate SI unit, and an appropriate number of significant figures

Appendix 4

A sample assessment activity and assessment schedule for PUS 6380

The following level 2 assessment activity from the *Assessment Guide: Physics* illustrates how students can be assessed against Unit Standard 6380 and is set out using the NZQA recommended format. In reading this activity reference should be made to elements 1 and 2 of Unit Standard 6380. A copy of this Unit Standards is reproduced in appendix 3.

SHUNTING

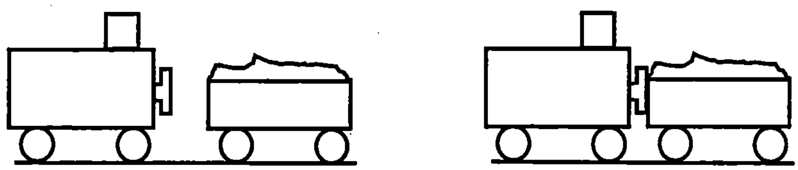
This activity assesses:

Unit:	6380	Apply formulae, graphical and vectorial methods to find unknowns for a physical system (Level 2).
Element:	1	Recognise the principles and quantities of a physical system.
	2	Determine an unknown quantity using a formula.

CONDITIONS: Formal test

INSTRUCTIONS:

A railway wagon of mass 2.0×10^3 kg moving at a 8.0 ms^{-1} collides with a 4.0×10^3 kg wagon which is stationary. As they collide, the wagons couple and move off together.



a) Identify the quantity that is conserved in this collision.

1.1❑

b) State a formula that can be used to calculate the velocity of the wagons immediately after the collision. Clearly state the symbol, value and unit of all variables used in the formula.

____

____

____

2.1☐
1.2☐

c) Calculate the velocity of the wagons after the collision. Show full working and give your answer as a full sentence complete with SI unit and an appropriate number of significant figures.

Substitution____

Sentence answer and unit____

____

2.2☐
2.3☐

Assessment Schedule: Shunting

Unit Standard 6380

Task Number	Element and Performance Criterion	Evidence	Judgement
a)	1.1	Using the principle of conservation of momentum.	The principle is correctly identified
b)	1.2	M_1 = mass of moving wagon = 2000 kg M_2 = mass of stationary wagon = 4000 kg V_1 = speed of moving wagon before collision = 8.0 m s^{-1} V_f = final speed of two wagons.	Symbols are correctly assigned with physical quantities, values and SI units.
	2.1	$M_1V_1 = (M_1 + M_2)\,V_f$	A correct formula consistent with conservation of momentum is shown.
c)	2.1	2000 x 8 = (2000 + 4000)V_f $V_f = \dfrac{16000}{6000}$	The calculation shows correct substitution.
	2.2	= 2.7 m s^{-1}	Correct answer.
	2.3	The speed of the two wagons after the collision is 2.7 m s^{-1}	Answer is given in a complete sentence. Numerical answer is correct to two significant figures.

Appendix 5

Moderation Action Plan for the Level 1-3 PUS

Contact **Support Officer, Moderation Services**
Science
New Zealand Qualifications Authority
P O Box 160
Wellington
Telephone: (04) 802 3000
Facsimile: (04) 802 3113

Moderation Option

A centrally established and directed external moderation system has been set up by NZQA on behalf of the Science and Technology Advisory Group.

MODERATION FOR LEVELS 1-3

a *Moderation System*
The New Zealand Qualifications Authority will be responsible for the implementation, operation, and monitoring of this moderation system.

The moderation action plan for science has been divided into two parts:

- Moderation Process
- Organisation of the Moderation Process

b *Moderation Process*
Moderation is a quality assurance process. In combination with other quality assurance processes such as registration of standards, accreditation, and audit, its aim is to ensure valid, fair and consistent assessment decisions.

Research has shown that effective moderation involves our interrelated procedures:

- Signalling standards
- Raising assessor expertise
- Ensuring the comparability of assessment activities and assessment schedules before they are used
- Verifying assessor decisions

These four procedures are central to the moderation process for science.

c *Signalling Standards and Raising Assessor Expertise*

Although unit standards provide the basic definition of the standards to be assessed it is essential that assessors receive clear advice and guidance on the interpretation of the unit standards. An assessment guide specifically related to the unit standards has been developed to signal standards.

Moderators with expertise in science assessment also have a role in signalling standards and raising the expertise of assessors. Moderation of assessment material will provide practical opportunities for advising assessors on the required standard.

d *Pre-assessment Moderation*

Every provider assessing unit standards at each level will be required to submit to a moderator their assessment programme, including the unit standards, which will be assessed, and a brief description of the assessment activities, which will be used.

Each year one or two unit standards in each domain (i.e. Biology, Chemistry, Physics, Science – Core, Earth Science), will be nominated by the national moderators for moderation. All unit standards assessed against by a provider, at each level, will be moderated at least once in every four years.

In response to individual providers assessment programmes a modified moderation schedule may be negotiated in advance with the local moderator. Once the unit standards for moderation have been agreed between the local moderator and the provider the following moderation process will occur. For each unit standard for moderation, the provider submits to a moderator for evaluation the assessment activities and schedules that they intend to use. If the moderator requires any modifications or changes to the assessment activities and/or assessment schedules these must be implemented and approved before assessment can take place.

Where the moderator has concerns about the ability of a provider to interpret standards consistently, additional assessment activities and assessment schedules may be requested.

Assessment standards must be adjusted before assessment occurs as it is not possible to adjust assessment decisions after assessment has taken place.

e) *Post-assessment Review and Verification*

Although assessment schedules and activities will be evaluated before they are undertaken, it is essential that the assessment decisions made from these assessments by providers for the award of credit are reviewed and verified. This provides a further check on the comparability of providers' assessments.

For the unit standards required to be moderated each year, the moderator will request randomly selected samples of student performance to be sent to the moderator for review and subsequent verification. The sample requested from each provider for each unit standard will include:

- 3 students who have achieved just above the credit boundary
- 3 students who have achieved just below the credit boundary.

The moderator may ask for additional samples and may even require samples from other unit standards to be submitted.

This local monitoring of assessor decision making is also supported by a nationally prescribed assessment activity (NPA), set nationally, and administered by providers under uniform conditions. It will be assessed by the providers according to national assessment schedules, and a sample of these assessments is sent to the moderators for review. Information gained from the NPA is used as part of the verification process. Where the moderators identify inconsistencies in assessment decisions between the NPA and the provider-based assessments, the differences will be discussed with the provider. The information gained from the NPA can be used by providers as part of their evidence for assessment for the award of credit. It may also be used as a reference point by the Qualifications Authority for national monitoring of provider standards to demonstrate that comparable standards are being maintained.

f) *Organisation of the Moderation Process*

Each year a National Moderator will be trained and appointed by NZQA Moderation Services to oversee the operation of the moderation process and to coordinate a national network of regional and local moderators. The national moderator will be responsible for monitoring the work of the regional and local moderators. The national moderator will also be responsible for the development of materials to facilitate and support the moderation process.

There will be a national network of local moderators each responsible for the moderation of up to ten providers. These moderators will be trained by the national moderator with support from the Qualifications Authority. The role of the local moderator will be to ensure the comparability of their providers' assessments activities and assessment schedules and to review and verify the assessment decisions made by these providers.

The local moderator will, one a year, facilitate a one-day local provider group meeting at a date mutually convenient to all local providers. One representative from each provider will be invited to attend the group meetings. Attendance will be voluntary. The main purpose of the local group meeting will be assessor development. The group will be led by the local moderator who will have the responsibility of facilitating understanding of national standards as conveyed through unit standards, exemplars and samples of learner performance.

The local moderators will be provided with documentation that will enable them to keep records of their moderation activities. They will be required to send regular reports, together with samples of their work, to their regional moderator. The regional moderator will be required to monitor the work of the local moderators and to follow up any irregularities that occur. The regional moderators will submit regular written reports to the national moderator.

There will also be regional meetings of local moderators. These meetings will review local interpretations of standards and provide a forum for the resolution of any local problems.

All the problems unsolved locally will be resolved either at regional meetings or by the national moderator when timing renders it impossible to have them discussed at a regional meeting. Regional meetings will also be a mechanism for providing feedback on the quality of the unit standards through the national moderator to the Science and Technology Advisory Group.

The national moderator will be present at some regional meetings and will thereby ensure consistency between the regional groups. This will be reinforced through annual meetings of the regional moderators.

Non-compliance with moderation requirements will initially be dealt with by the local moderator. Continued non-compliance will be immediately referred to the national moderator. If the national moderator cannot resolve the problem the matter would be referred to NZQA Moderation Services who will deal with any non-compliance in accordance with NZQA's non-compliance policy.

The national moderator will prepare an annual report, which will be presented to the New Zealand Qualifications Authority. NZQA Moderation

Services will produce an annual report (incorporating the national moderator's report) to the Science and Technology National Standards Body on the administration of the moderation system. These annual reports will be made available to providers.

Appendix 6

1996 Physics Regional and Local Moderator Questionnaire.

John Boereboom
Christchurch College of Education
PO Box 31 065
Christchurch

Dear fellow moderator

Over the next three years I will be conducting some research into various aspects of the moderation process for the Physics Unit Standards. This research will contribute towards my PhD and is not undertaken as part of my role as National Moderator.

I am particularly interested in the extent to which the moderation process is achieving comparability between schools and consistency between moderators. I also want to document any changes which occur over time as providers and moderators become more experienced.

Would you please:

- Moderate the attached assessment activity and schedule and record your moderation decisions on form PHYAMO2. The assessment activity to be moderated is deliberately flawed for the purpose of this research
- Complete the attached questionnaire.
- Return the questionnaire and moderated activity to me by 30 August 1996.

The data you provide will remain confidential. The analysis will be used for statistical purposes only and individuals will not be identified in the report.

Since this activity is not initiated by NZQA the time spent on moderating this activity can not be claimed on your moderator log.

Yours sincerely

John Boereboom
Senior Physics Lecturer

1996 Physics Regional and Local Moderator Questionnaire

Name: ____

Please complete the questionnaire by either circling the appropriate answer or by writing your response in the space provided.

1. Are you a local or a regional moderator?

   Local   Regional

2. How many years secondary teaching experience have you had?

   ___________ years.

3. How many years experience have you had of teaching Sixth Form physics

   ___________ years.

4. What is the highest level of physics training in your degree?

   Year 1   Year 2   Year 3   Honours   Masters   PhD

   Other:____________________

5. How successful was the moderator training you received in helping you to develop a clear view of your role as a local or regional moderator?

Very successful	Successful	Not sure	Unsuccessful	Very unsuccessful
5	4	3	2	1

Comment:

____

____

____

6. How successful was the moderator training you received in helping you to develop an understanding of the Physics Moderation Action Plan?

Very successful	Successful	Not sure	Unsuccessful	Very unsuccessful
5	4	3	2	1

Comment:

____

____

____

7. How successful was the moderator training you received in helping you to develop an understanding of how to carry out moderation of assessment activities and schedules?

Very successful	Successful	Not sure	Unsuccessful	Very unsuccessful
5	4	3	2	1

Comment:

____

____

____

8. How satisfactory do you think each of the following aspects of the physics moderation system is in achieving comparability between schools?

(a) Moderation of assessment activities

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

____

____

____

(b) Moderation of assessor judgements

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

____

____

____

(c) Communication with your Regional Moderator.

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

____

____

____

(d) Communication from the National Moderator.

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

____

____

____

(e) Communication with the contact person in your allocated provider schools.

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

____

____

____

9. How would you describe the handling of your moderator workload in addition to your teaching?

Very difficult	Difficult	Average	Easy	Very easy
5	4	3	2	1

Comment:

____

____

____

10. On average how many minutes does it take you to moderate an assessment activity?

_______________ minutes

11. What percentage of assessment activities you moderate would you approve for immediate use?

_______________ %

12. What do you see as the main strengths of the moderation system as it is operating at present?

____

____

____

13. What do you see as the main weaknesses of the moderation system as it is operating at present?

____

____

____

14. Please describe briefly any issues you have encountered in your dealings with schools. (You may continue on the back if necessary)

____

____

____

15. How many minutes did it take you to moderate the attached assessment activity and schedule?

______________ minutes

Thank you for taking the time to complete this questionnaire.

Appendix 7

1997 Physics Regional and Local Moderator Questionnaire

Name: ____

Please complete the questionnaire by either circling the appropriate answer or by writing your response in the space provided. Section A is to be answered by all moderators. Section B is to be answered only by moderators who commenced their duties this year.

Section A

To be answered by all moderators.

1. Are you a local or a regional moderator?

   Local   Regional

2. How many of the providers you were allocated are currently assessing against the physics Unit Standards?

   ________________ providers

3. How satisfactory do you think each of the following aspects of the physics moderation system is in achieving comparability between schools?

(a) Moderation of the assessment plan.

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

____

____

____

(b) Moderation of assessment activities.

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

(c) Moderation of assessor judgements

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

(d) The consistency of moderator decisions between moderators

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

(e) The check moderation process.

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

____

____

____

(f) Communication with your Regional Moderator.

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

____

____

____

(g) Communication from the National Moderator.

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

____

____

____

(h) Communication with the contact person in your allocated provider schools.

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

4. What do you consider to be the role of the NPA in the moderation process?

5. How necessary do you think the NPA is to the moderation process?

Very necessary	Necessary	Not sure	Not necessary	Very unnecessary
5	4	3	2	1

Comment:

6. When during the year should the NPA be held?

(Please circle the option you prefer)

During an NZQA specified window period

At a naturally occurring time in a providers assessment programme

Comment:

7. The NPA results from each school should be used as a statistical check on the way schools award credit

Stongly agree	Agree	Average	Disagree	Strongly disagree
5	4	3	2	1

Comment:

8. How would you describe the handling of your moderator workload in addition to your teaching?

Very difficult	Difficult	Average	Easy	Very easy
5	4	3	2	1

Comment:

9. On average how many minutes does it take you to moderate an assessment activity?

_______________ minutes

10. What percentage the of assessment activities you moderate would you approve for immediate use i.e. do not require resubmission?

_______________ %

11. What do you see as the main ***strengths*** of the moderation system as it is operating at present?

____

____

____

12. What do you see as the main ***weaknesses*** of the moderation system as it is operating at present?

____

____

____

13. Please describe briefly any issues you have encountered in your dealings with schools in your role as a moderator.

____

____

____

Section B

This section is to be answered only by moderators who commenced their duties in 1997.

14. How many years' secondary teaching experience have you had?

___________ years.

15. How many years' experience have you had of teaching Sixth Form physics

___________ years.

16. What is the highest level of physics training in your degree?

Year 1	Year 2	Year 3	Honours	Masters	PhD

Other:____________________

17. How successful was the moderator training you received in helping you to develop:

(a) a clear view of your role as a local or regional moderator?

Very successful	Successful	Not sure	Unsuccessful	Very unsuccessful
5	4	3	2	1

Comment:

____

____

____

(b) an understanding of the Physics Moderation Action Plan?

Very successful	Successful	Not sure	Unsuccessful	Very unsuccessful
5	4	3	2	1

Comment:

____

____

____

(c) an understanding of how to carry out moderation of assessment activities and schedules?

Very successful	Successful	Not sure	Unsuccessful	Very unsuccessful
5	4	3	2	1

Comment:

____

____

____

Thank you for taking the time to complete this questionnaire.

Appendix 8

1998 Physics Regional and Local Moderator Questionnaire

Name: ____

Please complete the questionnaire by either circling the appropriate answer or by writing your response in the space provided. Section A is to be answered by all moderators. Section B is to be answered only by moderators who commenced their duties this year.

Section A

To be answered by all moderators.

1. a) Are you a local or a regional moderator?

   Local   Regional

   b) How many years have you been a physics moderator?

   0   1   2   3

2. How many of the providers you were allocated are currently assessing against the physics Unit Standards?

   ________________ providers

3. Are you satisfied with the geographical distribution of the cluster of your allocated providers providers?

   YES   NO

Comment:

____

____

____

4. How satisfactory do you think each of the following aspects of the physics moderation system is in achieving comparability between schools?

(a) Moderation of the assessment plan.

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

____

____

____

____

(b) Moderation of assessment activities.

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

____

____

____

____

(c) Moderation of assessor judgements.

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

____

____

____

____

(d) The consistency of moderator decisions between moderators.

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

____

____

____

____

(e) The check moderation process.

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

____

____

____

____

(f) Communication with your Regional Moderator.

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

____

____

____

____

(g) Communication with the National Moderator.

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

(h) Communication with the contact person in your allocated provider schools.

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

(i) Meetings between Local Moderators and Providers.

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

(j) Meetings between Regional and Local Moderators.

Very satisfactory	Satisfactory	Average	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

---

5. Do you think meetings of Local Moderators with providers are necessary

YES NO

Comment:

---

6. When during the school year is the best time to hold:

a) meetings of Local Moderators with providers?

Middle Early Late

Comment:

---

b) meetings of Regional Moderators with Local Moderators?

Middle Early Late

Comment:

---

7. How important do you think it is that reassessment activities are moderated?

Very important	Important	Not sure	Unimportant	Very unimportant
5	4	3	2	1

Comment:

____

____

8. Do you agree with the decision not to have an NPA this year?

YES NO

Comment:

____

____

____

____

9. Do you agree with the decision to have pre-moderated activities available for providers to use?

YES NO

Comment:

____

____

10. In your opinion should the moderation process require evidence of authenticity of student work which is completed under informal conditions and submitted for verification?

YES NO

Comment:

____

11. In your opinion, how many items of marked student work are sufficient for the local moderator to verify assessor judgements?

8	6 (present case)	4	2

Comment:

____

____

12. In your opinion, how many times does a student need to demonstrate that they have mastered the learning outcomes of an element before they can receive credit?

3	2	1

Comment:

____

____

____

13. How satisfied are you that the conditions under which students are assessed are comparable between schools.

Very satisfied	Satisfied	Average	Unsatisfied	Very unsatisfied
5	4	3	2	1

Comment:

____

____

14. How consistent do you think providers are in their interpretation of when students can resubmit work.

Very consistent	Consistent	Average	Inconsistent	Very inconsistent
5	4	3	2	1

Comment:

____

____

____

____

15. How would you describe the handling of your moderator workload in addition to your teaching?

Very difficult	Difficult	Average	Easy	Very easy
5	4	3	2	1

Comment:

____

____

____

16. On average how many minutes does it take you to moderate an assessment activity?

_______________ minutes

17. On average how many minutes does it take you to verify the marking of the six items of student work?

_______________ minutes

18. About what percentage of the assessment activities you moderate would you approve for immediate use i.e. do not require resubmission?

_______________ %

19. What do you see as the main ***strengths*** of the moderation system as it is operating at present?

____

____

____

20. What do you see as the main ***weaknesses*** of the moderation system as it is operating at present?

____

____

____

21. Please state any suggestions you may have for improving the physics moderation system or for making moderation requirements easier for providers.

____

____

____

22. Please describe briefly any issues you have encountered in your dealings with schools in your role as a moderator.

____

____

____

Section B

This section is to be answered only by moderators who commenced their duties in 1998.

23. How many years secondary teaching experience have you had?

____________ years.

24. How many years experience have you had of teaching Sixth Form physics?

____________ years.

25. What is the highest level of physics training in your degree?

Year 1 Year 2 Year 3 Honours Masters PhD

Other:____________________

26. How successful was the moderator training you received in helping you to develop:

(a) a clear view of your role as moderator?

Very successful	Successful	Not sure	Unsuccessful	Very unsuccessful
5	4	3	2	1

Comment:

____

____

____

(b) an understanding of the Physics Moderation Action Plan?

Very successful	Successful	Not sure	Unsuccessful	Very unsuccessful
5	4	3	2	1

Comment:

____

____

____

(c) an understanding of how to carry out moderation of assessment activities and schedules?

Very successful	Successful	Not sure	Unsuccessful	Very unsuccessful
5	4	3	2	1

Comment:

____

____

Thank you for taking the time to complete this questionnaire.

Appendix 9

Level 2 assessment activity for the 1996 moderator agreement trial

Please use the attached form (PHYAMO2) to moderate the following activity. The activity was designed to cover all of Unit Standard 6379, and contains some deliberate flaws.

Unit 6379: Energy, Momentum and Equilibrium.

1.a) State the conditions that must be met for momentum to be conserved in explosions and collisions. (1.1)

  b) Give two examples of situations where momentum would not be conserved. (1.1)

2. State Newton's Third Law of Motion. (1.2)

3. Classify the following collisions as elastic or inelastic. (1.3)

   a) A car collides with a tree and comes to rest.

   b) Two billiard balls collide on a rough surface.

   c) Two astronauts collide during a space walk in outer space

4. A rifle of mass 4.0 kg fires a bullet of mass 25 g at a speed of 407 ms^{-1}. What is the recoil speed of the rifle? (2.1,2.3)

5. A mass of 0.5 kg hung from the end of a spring extends the spring by 25 cm.

   a) Calculate the spring constant. (3.1,3.2,3.3)

   b) How much elastic potential energy is stored in the spring? (3.1,3.2,3.3)

6. A metre rule of negligible mass is pivoted at the 20 cm mark. A weight of 15 N is suspended from the 80 cm mark. Calculate the weight that must be suspended from the 10 cm mark to balance the ruler.

Assessment Schedule

Assessment schedule: Energy, Momentum and Equilibrium (Unit Standard 6379)			
Task no	Element & PC's	Evidence	Judgement statements
1a	1.1*	There must be no unbalanced forces acting.	Statement must contain the words unbalanced and external.
1b	1.1*	Situation in which gravity or friction acts.	Any two examples involving an external force.
2	1.2	To every action there is an equal and opposite reaction.	Or similar statement.
3a	1.3*	Inelastic	Must have correct choice.
3b	1.3*	Elastic	Must have correct choice.
3c	1.3*	Elastic	Must have correct choice.
4	2.1 2.2 2.3	$M_1 U_1 + M_2 U_2 = M_1 V_1 + M_2 V_2$ $0 = - 0.4 \times U_1 + 0.025 \times 407$ $U_1 = 3.54$ m s^{-1} The gun recoils at 3.5 m s^{-1}	Correct formula selected. Values substituted correctly. Answer is stated as a sentence and includes correct SI unit.
5a	3.1* 3.2* 3.3*	$k = F / x$ $k = mg / x$ $k = 0.5 \times 10 / 0.25$ $k = 20$ N m^{-1} The spring constant is 20 N m^{-1}	Correct formula selected. Values substituted correctly. Answer is stated as a sentence and includes correct SI unit.
5b	3.1* 3.2* 3.3*	$E_{elastic} = 1/2\ k x^2$ $E_{elastic} = 10 \times 0.25^2$ $E_{elastic} = 0.625$ J The elastic potential energy stored in	Correct formula selected. Values substituted correctly.

		the spring is 0.625 J	Answer is stated as a sentence and includes correct SI unit.
6	4.1 4.2 4.3	The sum of the clockwise moments about the pivot is equal to the sum of the anticlockwise moments about the pivot. 15 x 0.6 = W x 0.1 W = 90 N The ruler will balance when a 90 N weight acts down on the 10 cm mark.	Correct formula selected. Values substituted correctly. Answer is stated as a sentence and includes correct SI unit.
* performance criteria must be met on all occasions indicated for credit of the element			

Moderation of sample Energy, Momentum and Equilibrium assessment activity by the National Moderator. (Unused space on the form has been deleted)

PHYAMO2

PHYSICS UNIT STANDARD MODERATION

Assessment Material Level 1, ②, 3 (Circle the level that applies)

Provider: *Check Moderation*

Activity: *Energy, Momentum and Equilibrium.*

Unit Standard No: *6379*

Assessment Activity

	Yes	No
Does the activity state the Unit Standard number, Title and element to be assessed?		X
Is the language appropriate?	X	
Are the instructions clear?		X

Comment on the modifications required:

The activity does not state the Unit Standard title nor the elements to be assessed.

	Yes	No
Is the activity at an appropriate standard?	X	

Comment on the modifications required:

	Yes	No
Do the instructions give the student the opportunity to meet the requirements on the unit Standard(s) and selected element(s), i.e. is the activity valid?		X
Are the special notes being adhered to	X	
Range statements covered?		X

Comment on the modifications required:

In question 2, the requirement to state Newton's third law is insufficient. To meet p.c. 1.2 students have to describe an actual situation in terms of the forces acting. An extra part to this question which requires students to label a physical situation must be added. Only two parts of the range statement for element 2 are covered in question 4, there is no coverage of the mass and kinetic energy component of the range statement.

Comments to providers if necessary:

Add a further question to get students to calculate mass and kinetic energy in a one dimensional collision

Assessment Schedule

	Yes	No
Is there a clear link between the selected elements of the Unit Standard (s) and required evidence?	X	
Are the evidence statements consistent with the requirements of the Unit Standard?		X
Do the evidence statements indicate a range of possible answers expected from the learner?	X	

Comment on the modifications required:

The answer to part b is wrong. The correct answer is " inelastic". The answer to part 4 is wrong . The correct answer is - 2.5 m s^{-2}.

	Yes	No
Do the judgement statements clearly describe acceptable levels of performance(quality and quantity)?	X	

Comment on the modifications required:

It is possible to meet p.c.'s 2.3, 3.3, and 4.3 without stating the answer in the form of a sentence. The alternative answers must be clearly signalled in the judgement statements i.e. a mathematical sentence is adequate.

Tick One:

Material acceptable for use: ☐

Material acceptable for use with the above modifications: ☐

Resubmit material: ☒

Specify material to be resubmitted____________________

By (date) ___________

General comments to provider:

Make changes as recommended and resubmit the activity.

Date sample of assessed work due to moderator: N.A.

SIGNATURE OF MODERATOR: *John Boereboom* **DATE:** *23 June 1996*

Summary of moderators' decisions and level of consistency on moderation criteria for the 1996 sample Level 2 assessment activity

Moderation criteria applied to the 1996 *Energy, Momentum and Equilibrium* assessment activity for Level 2 US 6379.	**Yes (%)**	**No (%)**
Does the activity state the Unit Standard number, title and elements to be assessed?	31	69*
Is the language appropriate?	69*	31
Are the instructions clear?	38	62*
Is the activity at an appropriate standard?	85*	15
Do the instructions give the students the opportunity to meet the requirements of the Unit Standard and selected element(s). ie is the activity valid?	15	85*
Are the special notes being adhered to?	62*	38
Are the range statements covered?	15	85*
Is there a clear link between the selected elements of the Unit Standard(s) and required evidence?	69*	31
Are the evidence statements consistent with the requirements of the Unit Standard?	31	69*
Do the evidence statements indicate a range of possible answers to be expected from the learner?	54*	46
Do the judgement statements clearly describe acceptable performance levels? (quality and quantity)	62*	38
Material acceptable for use	8 %	
Material acceptable for use with the above modifications	23%	
Resubmit material	69%*	
Average percentage of moderator agreement on all criteria (SD)	70 (10)	

* National Moderator's decisions

Appendix 10

Level 2 assessment activity for the 1997 moderator agreement trial

Please moderate this activity using Form SCI02. Keep a record of the amount of time it takes to complete the moderation.

VERTICAL OSCILLATIONS OF A SPRING

This activity assesses:

Unit:	6387	Use graphical analysis to determine simple non-linear relationships (Level 2).
Element:	1	Graph data.

CONDITIONS: Formal class test

INSTRUCTIONS:

The period is measured for several different masses oscillating on a spring. The results are in the table below.

M (kg)	0	0.064	0.26	0.58	1.0	1.6	2.3	3.1
T (s)	0	0.4	0.8	1.2	2.4	2.0	2.4	1.8

a) Use the data to plot an appropriate graph.

1 1❑
1.2❑
1.3❑
1 4❑

b) Use the shape of the graph to predict the relationship between M and T. State the predicted relationship.

2.1❑

c) In order to obtain a new graph that is a straight line, the data needs to be processed. Generate a third column of processed data with a suitable heading.

d) Using your new data, plot a second graph to verify the predicted relationship.

2.2❑

e) Determine the gradient and Y intercepts of the second graph.

2.3❑

f) State the mathematical relationship between M and T.

2.4❑

g) Use the linear graph to find what mass needs to be added to the spring to get a period of 1.4 s. Show construction lines on your graph. 2.5❑

Assessment Schedule: Vertical oscillations of a spring

Unit Standard 6386

Task Number	Element and Performance Criteria	Evidence (The answers or performance expected from the students)	Judgement (A statement that defines the standard to be achieved)
a)	1.1	Student graph.	Period on Y axis, Mass on X axis.
	1.2	Student graph.	Axis labelled with quantity.
	1.3	Student graph.	Points plotted correctly.
	1.4	Student graph.	Line drawn correctly.
b)	2.1	Mass is proportional to the square of the period, orT α M^2.	Square relationship is implied.
c)		Completed third column headed T^2 in table of results.	Not assessed.
d)	2.2	Linear graph drawn.	Points plotted are consistent with processed data. Best fit line drawn.
e)	2.3	Gradient calculated from graph = 0.4 y = 0	No units required. Range for gradient is 0.35 to 0.45.
f)	2.5	$M = 0.4\ T^2$	Accept any mathematically equivalent statement.
g)	2.5	M = 0.78 kg	Accept m in the range 0.76 to 0.80.

It took me ___________ minutes to moderate this activity.

The activity and assessment schedule had some deliberate flaws inserted.

The deficiencies in the activity are:

- the Unit Standard number is wrongly stated as 6387 it should be 6386
- Element 2 is missing from the description of which elements the activity assesses against
- the conditions for administering the activity are not stated

- the values in the table are not consistent with the relationship between M and T
- the number of decimal places in the readings are not consistent

The deficiencies in the assessment schedule are:

- a graph of student results should be included:
- the judgement statement for performance criteria 1.2 should include the requirement of units
- the line drawn for task (a) should be a line of best fit
- the evidence statement for task b should read M α T^2
- task (c) should refer to performance criteria 2.2
- Task (f) assesses against performance criteria 2.4, not 2.5 as stated

Summary of moderators' decisions and level of consistency on moderation criteria for the 1997 sample Level 2 assessment activity

Moderation criteria applied to the 1997 Level 2 assessment activity. *Vertical Oscillations of a Spring* for US 6386.	**Yes (%)**	**No (%)**
Are the Unit Standard number, title and selected element(s) to be assessed given?	5	95*
Are the instructions clear and easily understood?	53*	47
Is the level of difficulty appropriate?	100 *	0
Does it give the learner the opportunity to meet the all the requirements of the selected element(s), i.e. is the activity valid?	68*	32
Are the evidence statements consistent with the requirements of the Unit Standard?	11	89*
Do evidence statements indicate a range of possible expected learner responses?	26	74*
Do judgement statements clearly describe acceptable performance levels,(quality and / or quantity)?	21	79*
Material acceptable for use	0%	
Material acceptable for use with the above modifications	32%	
Resubmit material	68%*	
Average percentage of moderator agreement on all criteria (SD)	80 (16)	

* National Moderator's decisions.

Appendix 11

Level 2 assessment activity for the 1998 moderator agreement trial

Instructions: Please moderate this activity using the attached Form Mod 02. Keep a record of the amount of time it takes to complete the moderation.

WAVES

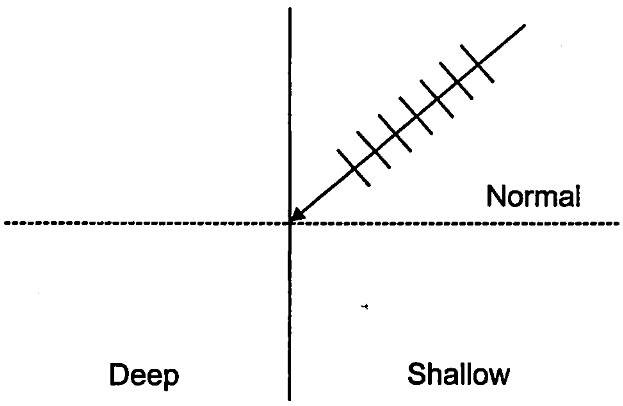
This activity assesses:

Unit: 6383 Demonstrate knowledge of waves (Level 2).
Element: 1 Describe waves.

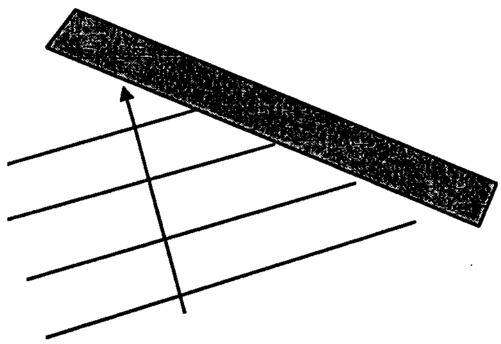
CONDITIONS: Formal class test

Task 1

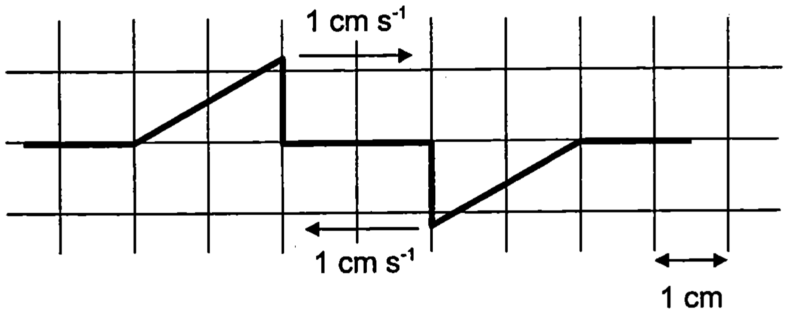
a) The diagram shows waves crossing from shallow to deep water. Complete the diagram with an arrow to indicate the direction of water movement in the deep water.



b) The diagram shows a wave hitting a solid border. Draw the reflected wave carefully, showing its direction of travel. 1.1❑



c) The diagram shows two pulses approaching each other on a string, both moving at 1 cm s^{-1}. Complete the diagram to show the shape of the string two seconds later, ie when they completely overlap.



1.1❑

The speed of sound in air is 340 m s^{-1}. A particular sound wave has a frequency of 550 Hz.

d) Calculate the wavelength in air of this sound wave.

Formula and substitution____

____

____2.1❑

Answer sentence and unit____

____2.2❑

e) The wavelength of this sound wave in water is 2.7 m. Calculate the speed of sound in water.

Formula and substitution____

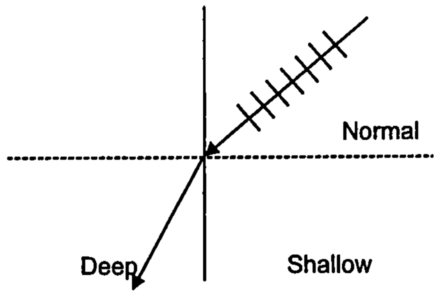
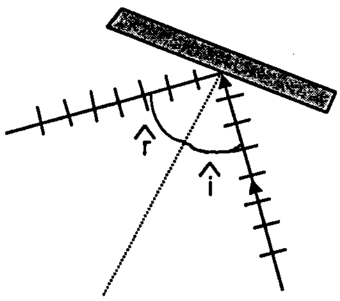
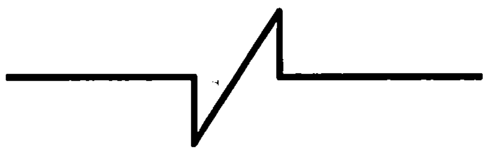
____

____2.1❑

Answer sentence and unit____

____2.2❑

Assessment Schedule: Waves
Unit Standard 6382

Task Number	Element and Performance Criteria	Evidence (The answers or performance expected from the students)	Judgement (A statement that defines the standard to be achieved)
1 a)	1.1		Wave refracts away from the normal.
1 b)	1.1		Wave fronts reflected such that angle i= angle r.
1 c)	1.2		Displacements combined correctly.
1 d)	2.1*	$v = f\lambda$ $\lambda = 340/550$	Correct formula and substitution.
	2.2*	$\lambda = 0.62$ m	Correct answer and unit.
1 e)	2.1*	$v = f\lambda$ $v = 550 \times 0.62$	Correct formula and substitution.
	2.2*	$v = 1500 \text{ m s}^{-1}$	Correct answer and unit.

* performance criteria 2.1 and 2.2 may be met in either 1(d) ***or*** 1(e)

It took me ____________ minutes to moderate this activity.

Points to note

Assessment activity

1. The unit standard number is wrongly stated as 6383, it should be 6382
2. Element 2 is not included in the introductory information even though the activity assesses against it.
3. The performance criterion assessed against in part 1 (c) is 1.2, not 1.1 as indicated by the tick box.

Assessment schedule

1. The judgement statement for performance criterion 1.1 should state that the wave refracts towards the normal not away from the normal.
2. In part 1(e) the wavelength is incorrectly stated as 0.62 m, it should be 2.7 m.
3. The footnote to the assessment schedule incorrectly specifies that performance criteria 2.1 and 2.2 may be met in either 1(d) ***or*** 1(e). This should read " performance criteria 2.1 and 2.2 must be met in 1(d) ***and*** 1(e) for credit of element 1.

Summary of moderators' decisions and level of consistency on moderation criteria for the 1998 sample Level 2 assessment activity

Moderation criteria applied to the 1998, Level 2 assessment activity *Waves* for US 6382, *Demonstrate knowledge of waves.*	**Yes (%)**	**No (%)**
Are the Unit Standard number, title and selected element(s) to be assessed given?	0	100*
Are the instructions clear and easily understood?	68*	32
Is the level of difficulty appropriate?	100*	0
Does it give the learner the opportunity to meet the all the requirements of the selected element(s), i.e. is the activity valid?	11	89*
Are the evidence statements consistent with the requirements of the Unit Standard?	21	79*
Do evidence statements indicate a range of possible expected learner responses?	42	58*
Do judgement statements clearly describe acceptable performance levels,(quality and / or quantity)?	11	89*
Material acceptable for use	0	
Material acceptable for use with the above modifications	11	
Resubmit material	89*	
Average percentage of moderator agreement on all criteria (SD)	84 (15)	

* National Moderator's decisions

Appendix 12

1996 Level 2 Physics Unit Standards Questionnaire

This questionnaire is to be completed by the teacher in charge of physics. Please answer each of the following questions by circling the appropriate category or by writing a comment in the space provided.

1. Background information

Name:____

School:____

Type of school: State Private Integrated Boys Girls Co-educational

Number of students involved in the level two Physics Unit Standard trial:______________

2. The Physics Unit Standards

a) How well do you consider the level 2 Physics Unit Standards to reflect the Physics Curriculum?

Very well	Quite well	Not Sure	Poorly	Very poorly
5	4	3	2	1

b) How useful do you find the performance criteria within each element to indicate student mastery of the learning outcome stated in the element?

Very useful	Useful	Not Sure	Of limited use	No use
5	4	3	2	1

Comment:

____

____

c) How do you think your students feel about Unit Standards assessment in physics?

Very positive	Positive	Not Sure	Negative	Very negative
5	4	3	2	1

Comment:

d) How do you think assessing against the Physics Unit Standards has affected student motivation in your physics course?

Very favourably	Favourably	Not Sure	Unfavourably	Very unfavourably
5	4	3	2	1

Comment:

3. Curriculum coverage

The following list of topics is based on the curriculum level 7 content described in *Physics in the New Zealand Curriculum*. Please place a tick beside each topic that you have will have completed with your class by the end of this year.

Topic:	
Kinematic equations	
Relative motion	
Vectors	
Projectile motion	
Circular motion	
Levers and torque	
Momentum	
Impulse	
Elastic and inelastic collisions	
Hooke's law	
Elastic potential energy	
Parallel and series circuits	
Ohms law	
Voltage dividers	
Non-ohmic conductors	
Magnetic force on a current carrying conductor	
Magnetic force on a charged particle	
Electromagnetic induction	
Simple generator	
Reflection and refraction of light waves	
Reflection and refraction of water waves	
Superposition of waves	
The electromagnetic spectrum	
Atomic models	
Half life	
Radioactive decay	
alpha, beta and gamma emission	

4. Unit Standard Assessment

a) Please indicate the amount of class time devoted this year to teaching, assessing and reassessing of each of the following level 2 Unit Standards you have assessed against this year. If you did not offer the Unit Standard please write NA.

		Number of class periods spent:		
Unit no	Unit title	Teaching to prepare students for this Unit Standard	Assessing against this Unit Standard	Reassessing students for this Unit Standard
6378	Demonstrate knowledge of motion in one and two dimensions.			
6379	Demonstrate understanding of energy, momentum and equilibrium.			
6380	Apply Formulae, graphical and vectorial methods to find unknowns for a physical system.			
6381	Describe, construct and determine unknowns for electrical and electromagnetic systems.			
6382	Demonstrate knowledge of waves.			
6383	Describe the development of a selected physics idea and a physics-based application.			
6384	Investigate a physical system to determine a relationship with supervision.			
6385	Carry out a practical investigation of a physics-based application with supervision.			
6386	Use graphical analysis to determine simple non-linear physical relationships.			
6387	Demonstrate knowledge of elementary nuclear physics and radioactivity.			

b) What is the length of a class period at your school? minutes.

c) Please indicate which Unit Standards (if any) you think have an inappropriate number of credits allocated to them and explain why.

Unit Standard Number	**Reason:**

A valid course result is one that accurately describes student achievement relative to the course objectives.

d) How valid do you think Unit Standard credits are for indicating student achievement in Sixth Form Physics?

Very valid	Valid	Not Sure	Invalid	Very invalid
5	4	3	2	1

Comment:

____

____

____

5. Sixth Form Certificate Assessment

a) What is the total amount of class periods you have spent this year on assessment for Sixth Form Certificate? If you practised dual assessment include the time in both question 3 and 4. Include exams, tests, assessed practical work etc.

Class periods

b) How valid do you think a Sixth Form Certificate Grade is for indicating student achievement in Sixth Form Physics?

Very valid	Valid	Not Sure	Invalid	Very invalid
5	4	3	2	1

Comment:

____

____

____

6. Teacher Workload

State the average amount of time (in minutes) you spend each week on the following tasks associated with assessment and administration of the Level 2 Physics Unit standards. You may add extra tasks if necessary.

Task	Time (minutes)
Writing (re)assessment activities	
Marking	
Moderation	
Department meetings	
Other	

7. Moderation

a) What procedures are used within your physics department for ensuring that the Unit Standard assessment activities used for different level 2 physics classes are of a comparable standard?

____

____

____

b) What procedures are used within your physics department for ensuring that the assessor judgements of teachers of different level 2 physics classes are consistent?

____

____

____

c) How satisfactory do you think each of the following aspects of the physics moderation system is in achieving comparability between schools?

(1) Moderation of assessment activities

Very satisfactory	Satisfactory	Not Sure	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

____

____

____

(2) Moderation of assessor judgements

Very satisfactory	Satisfactory	Not Sure	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

____

____

____

(3) Communication with your local moderator

Very satisfactory	Satisfactory	Not Sure	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

____

____

____

d) How useful is the assessment Guide: Physics as an aid in establishing the difficulty level of assessment activities you write yourself?

Very useful	Useful	Not Sure	Of limited use	No use
5	4	3	2	1

Comment:

____

____

____

e) How useful did you find the cluster meetings during the trial of the level 2 Physics Unit Standards?

Very useful	Useful	Not Sure	Of limited use	No use
5	4	3	2	1

Comment:

____

____

____

f) How useful did you find the 3 days of training you received to prepare you for the introduction of the level 2 Physics Unit Standards?

Very useful	Useful	Not Sure	Of limited use	No use
5	4	3	2	1

Comment:

____

____

____

8. General

a) What have been the advantages of assessing against the Physics Unit Standards:

(1) for you as a teacher?

____

____

____

(2) for your students?

b) What have been the disadvantages of assessing against the Physics Unit Standards:

(1) for you as a teacher?

(2) for your students?

c) Have there been any organisational problems at your school specifically related to the trial?

Comment:

d) Which of the following best describes the effect of the Physics Unit Standards had on your classroom teaching?

Very positive	Positive	Not Sure	Negative	Very negative
5	4	3	2	1

Comment:

e) Which of the following best describes the effect of the Physics Unit Standards on student learning?

Very positive	Positive	Not Sure	Negative	Very negative
5	4	3	2	1

Comment:

f) Which of the following best describes your attitude towards the use of Unit Standards in Sixth Form Physics?

Very enthusiastic	Enthusiastic	Not Sure	Unenthusiastic	Very unenthusiastic
5	4	3	2	1

Please explain:

g) How useful was the communication you received from NZQA in relation to your schools participation in the level 2 Physics Unit Standards trial?

Very useful	Useful	Not Sure	Of limited use	No use
5	4	3	2	1

Comment:

h) Will your school assess against the Physics Unit Standards next year?

YES NO

If your answer was yes please indicate the levels at which the assessment will occur.

Level 2 Level 3

i) How did the PPTA freeze affect assessment against the Physics Unit Standards in your school?

____

____

____

j) Do you have any additional comments about the trial or the new assessment procedures?

____

____

____

Thank you for taking the time to complete this questionnaire. Please use the franked addressed envelope provided to mail your completed questionnaire and the completed student questionnaires to:

John Boereboom
Christchurch College of Education
PO box 31 065
Christchurch

Appendix 13

1996 Form 6 Physics Teacher Questionnaire

This questionnaire is to be completed by the teacher in charge of physics. Please answer each of the following questions by circling the appropriate category or by writing a comment in the space provided.

1. Background information

Name:____

School:____

Type of school: State Private Integrated Boys Girls Co-educational

Number of students taking Sixth Form Physics:______________________

2. Sixth Form Assessment

a) What is the total amount of class periods you have spent this year on assessment for Sixth Form Certificate? Include exams, tests, assessed practical work etc.

Class periods

b) What is the length of a class period at your school?

minutes.

A valid course result is one which accurately describes student achievement relative to the course objectives.

c) How valid do you think Unit Standard credits are for indicating student achievement in Sixth Form Physics?

Very valid	Valid	Not Sure	Invalid	Very invalid
5	4	3	2	1

Comment:

d) How valid do you think a Sixth Form Certificate Grade is for indicating student achievement in Sixth Form Physics?

Very valid	Valid	Not Sure	Invalid	Very invalid
5	4	3	2	1

Comment:

3. Curriculum coverage

The following list of topics is based on the curriculum level 7 content described in *Physics in the New Zealand Curriculum*. Please place a tick beside each topic that you have will have completed with your class by the end of this year.

Topic:	
Kinematic equations	
Relative motion	
Vectors	
Projectile motion	
Circular motion	
Levers and torque	
Momentum	
Impulse	
Elastic and inelastic collisions	
Hooke's law	
Elastic potential energy	
Parallel and series circuits	
Ohms law	
Voltage dividers	
Non-ohmic conductors	
Magnetic force on a current carrying conductor	
Magnetic force on a charged particle	
Electromagnetic induction	
Simple generator	

Reflection and refraction of light waves	
Reflection and refraction of water waves	
Superposition of waves	
The electromagnetic spectrum	
Atomic models	
Half life	
Radioactive decay	
alpha, beta and gamma emission	

4. Teacher Workload

State the average amount of time (in minutes) you spend each week on the following tasks associated with assessment for Sixth Form Certificate Physics. You may add extra tasks if necessary.

Task	Time (minutes)
Writing (re)assessment activities	
Marking	
Moderation	
Department meetings	
Maintaining student records	
Other	

5. Moderation

a) What procedures are used within your physics department for ensuring that the Unit Standard assessment activities used for different Sixth Form Physics classes are of a comparable standard

____

____

____

b) What procedures are used within your physics department for ensuring that the assessor judgements of teachers of different Sixth Form Physics classes are consistent?

____

____

____

c) Based on your observations od the trial schools, how satisfactory do you consider each of the following aspects of the physics moderation system to be in achieving comparability between schools?

(1) Moderation of assessment activities

Very satisfactory	Satisfactory	Not Sure	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

---

(2) Moderation of assessor judgements

Very satisfactory	Satisfactory	Not Sure	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:

---

6. General

a) What do you consider to be the main advantages (if any) of assessing against the Physics Unit Standards:

(1) for you as a teacher?

---

(2) for your students?

---

b) What do you consider to be the main disadvantages (if any) of assessing against the Physics Unit Standards:

(1) for you as a teacher?

(2) for your students?

c) How do you think your students feel about assessment in Sixth Form Physics

Very positive	Positive	Not Sure	Negative	Very negative
5	4	3	2	1

Comment:

d) What are your reasons for not taking part in the level 2 Physics Unit Standards trial this year?

e) Which of the following best describes your attitude towards the use of Unit Standards in Sixth Form Physics?

Very enthusiastic	Enthusiastic	Not Sure	Unenthusiastic	Very unenthusiastic
5	4	3	2	1

Please explain:

f) How useful was the communication you received from NZQA in relation to the level 2 Physics Unit Standards trial?

Very useful	Useful	Not Sure	Of limited use	No use
5	4	3	2	1

Comment:

___

___

___

g) Do you have any additional comments about the new assessment procedures?

___

___

___

Thank you for taking the time to complete this questionnaire. Please use the franked addressed envelope provided to mail your completed questionnaire and the completed student questionnaires to:

John Boereboom
Christchurch College of Education
PO box 31 065
Christchurch

Appendix 14

1997 Physics Assessment Questionnaire

This questionnaire is to be completed by the teacher in charge of physics.
The due date for the completion of the questionnaire is **Friday 13 June.**

Instructions

Please answer the questions by circling the appropriate category or by writing a comment in the space provided.If the question is not applicable please write N.A.

Section A is to be completed by all respondents.

Section B is to be answered only by teachers who are currently assessing against the Physics Unit Standards.

Section A

To be answered by all respondents

Background information

School:____

Type of school:	State	Boys
	Private	Girls
	Integrated	Co-educational

1. a) Are you assessing against the Physics Unit Standards this year?

   Yes No

   b) Did you assess against the Physics Unit Standards in 1996?

   Yes No

c) If you assessed against the Physics Unit Standards last year but are not this year, please state your main reason for discontinuing.

____

____

____

2. A valid course result is one which accurately describes student achievement relative to the course objectives.

a) How valid do you think a Sixth Form Certificate Grade is for indicating student achievement in Sixth Form Physics?

Very valid	Valid	Not sure	Invalid	Very invalid
5	4	3	2	1

Comment:____

____

____

____

b) How valid do you think Unit Standard credits are for indicating student achievement in Sixth Form Physics?

Very valid	Valid	Not sure	Invalid	Very invalid
5	4	3	2	1

Comment:____

____

____

____

3) Which of the following best describes your attitude towards the use of Unit Standards in Sixth Form Physics?

Very enthusiastic	Enthusiastic	Not sure	Unenthusiastic	Very unenthusiastic
5	4	3	2	1

Comment:____

____

____

____

4) Please indicate to what extent you agree or disagree with the following statements:

a) Unit standards are appropriate for assessing practical work skills.

Strongly agree	Agree	Not Sure	Disagree	Strongly disagree
5	4	3	2	1

Comment:____

b) Unit standards are appropriate for assessing students'abilty to solve physics problems.

Strongly agree	Agree	Not Sure	Disagree	Strongly disagree
5	4	3	2	1

Comment:____

c) Unit standards are appropriate for assessing students' ability to explain physics concepts.

Strongly agree	Agree	Not Sure	Disagree	Strongly disagree
5	4	3	2	1

Comment:____

d) Criteria for excellence should be built into each Unit Standard.

Strongly agree	Agree	Not Sure	Disagree	Strongly disagree
5	4	3	2	1

Comment:____

Curriculum coverage

5. The following list of topics is based on the curriculum level 7 content described in *Physics in the New Zealand Curriculum*. Please place a tick beside each topic you will have completed with your class by the end of this year.

Topic:	
Kinematic equations	
Relative motion	
Vectors	
Projectile motion	
Circular motion	
Levers and torque	
Momentum	
Impulse	
Elastic and inelastic collisions	
Hooke's law	
Elastic potential energy	
Parallel and series circuits	
Ohms law	
Voltage dividers	
Non-ohmic conductors	
Magnetic force on a current carrying conductor	
Magnetic force on a charged particle	
Electromagnetic induction	
Simple generator	
Reflection and refraction of light waves	
Reflection and refraction of water waves	
Superposition of waves	
The electromagnetic spectrum	
Atomic models	
Half life	
Radioactive decay	
Alpha, beta and gamma emission	

Assessment

6. For your sixth form physics class, what is the total number of hours of class time you will spent this year on assessment? Include exams, tests, assessed practical work etc. If not applicable, write NA in the box.

a) Sixth Form Certificate hours

b) Unit Standards hours

7. Estimate the average amount of time (in minutes) you spend each week on the following tasks associated with assessment and administration of sixth form assessment. You may add extra tasks if necessary. If not applicable, write NA in the box.

Task	Sixth Form Certificate Time (minutes)	Unit Standards Time (minutes)
Writing (re)assessment activities		
Marking		
Moderation		
Department meetings		
Other		

8. What procedures are used within your physics department for ensuring that the assessment activities used for different sixth form physics classes are of a comparable standard?

____

____

9. What procedures are used within your physics department for ensuring that the assessor judgements of teachers of different level 2 physics classes are consistent?

____

____

____

10. (a)Please rank the following options for physics assessment at the sixth form level in order of preference. Assign a 1 to your most preferred option and a 5 to you least preferred option.

Assessment option	**Rank**
National examination only	
Sixth Form Certificate (present model)	
Sixth Form Certificate (moderated using reference tests)	
Unit Standards only	
Sixth Form Certificate (present model) for problem solving and Unit Standards for practical work.	

Other (Please state):____________________________________

(b) Please rank the following options for physics assessment at the seventh form level in order of preference. Assign a 1 to your most preferred option and a 4 to you least preferred option.

Assessment option	**Rank**
National Bursary / Scholarship examination only	
National Bursary / Scholarship examination with an internally assessed component scaled to the exam.	
Unit Standards only	
A combination of examination and Unit Standards both of which contribute credit towards the National Certificate.	

Other (Please state):____________________________________

Section B

To be answered only by teachers who are currently assessing against the Physics Unit Standards

11. a) Please circle the level(s) at which the physics department in your school is currently assessing against the Physics Unit Standards.

Level 1 Level 2 Level 3

b) Please indicate the number of credits you will offer in your assessment programme at levels 2 and 3.

Level	Number of credits
2	
3	

12. Do you think the revised Unit Standards are an improvement on the Unit Standards used in the trial last year.

Definite improvement	Slight improvement	Not sure	No improvement	Worse
5	4	3	2	1

Comment:____

13. How well do you think the level 2 Physics Unit Standards reflect the Physics Curriculum?

Very well	Quite well	Not sure	Poorly	Very poorly
5	4	3	2	1

Comment:____

14. How useful do you find the performance criteria within each element for indicating student mastery of the learning outcome stated in the element?

Very useful	Useful	Not sure	Of limited use	No use
5	4	3	2	1

Comment:____

15. How do you think your students feel about Unit Standards assessment in physics?

Very positive	Positive	Not sure	Negative	Very negative
5	4	3	2	1

Comment:____

16. How do you think assessing against the Physics Unit Standards has affected student motivation in your physics course?

Very favourably	Favourably	Not sure	Unfavourably	Very unfavourably
5	4	3	2	1

Comment:____

Moderation

17. How satisfactory do you think each of the following aspects of the physics moderation system is in achieving comparability between schools?

a) Moderation of assessment activities

Very satisfactory	Satisfactory	Not sure	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:____

b) Moderation of assessor judgements

Very satisfactory	Satisfactory	Not sure	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:____

c) The Nationally Prescribed Activity

Very satisfactory	Satisfactory	Not sure	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:____

d) Communication with your local moderator

Very satisfactory	Satisfactory	Not sure	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:____

18. How useful are the revised assessment activities in the *Assessment Guide: Physics*?

Very useful	Useful	Not sure	Of limited use	No use
5	4	3	2	1

Comment:____

19. Which of the following best describes the effect assessing against the Physics Unit Standards has had on your classroom teaching?

Very positive	Positive	Not sure	Negative	Very negative
5	4	3	2	1

Comment:____

20. Which of the following best describes the effect of assessing against the Physics Unit Standards on student learning:

a) for the more able student?

Very positive	Positive	Not sure	Negative	Very negative
5	4	3	2	1

b) for the average student?

Very positive	Positive	Not sure	Negative	Very negative
5	4	3	2	1

c) for the less able student?

Very positive	Positive	Not sure	Negative	Very negative
5	4	3	2	1

Comment:____

21. Do you think assessment against the Physics Unit standards enables recognition of excellence?

Yes No

Comment:____

22. Please state any suggestions you may have about how criteria for recognising excellence can be built into assessment against Unit Standards.

____

____

____

23. How helpful was the communication you received from NZQA in explaining the administrative procedures associated with assessment and moderation of the Physics Unit Standards?

Very useful	Useful	Not sure	Of limited use	No use
5	4	3	2	1

Comment:______________________________________

____

____

____

24. What is your preferred method for receiving a copy of the Unit Standards form NZQA?

The c.d. Framework Explorer   A paper copy

Comment:______________________________________

____

____

____

25. a) Do you think there should be an NPA each year?

Yes   No

b) When should the NPA be available to administer to students?

Any time during the year

During a specified window period i.e. term

On specified dates

c) What is the purpose of the NPA?

26. Do you have any additional comments about assessment against Unit Standards, assessment in sixth form physics or implementation of the National Qualifications Framework? (Continue on the back of this page if necessary)

Comment:____

Thank you for taking part in this survey. Please return the completed questionnaire in the postage paid envelope provided by

John Boereboom
Christchurch College of Education
PO Box 31 065
Christchurch.

Appendix 15

1998 Physics Assessment Questionnaire

This questionnaire is to be completed by the teacher in charge of physics.
The due date for the return of the questionnaire is **Friday 29 May.**

Instructions

Please answer the questions by circling the appropriate category or by writing a comment in the space provided. If the question is not applicable please write N.A.

Section A is to be completed by all respondents.

Section B is to be answered only by teachers who are assessing against the physics unit standards in 1998.

Section A

To be answered by all respondents

Background information

School:____

Type of school:	State	Boys
	Private	Girls
	Integrated	Co-educational

1. a) Are you assessing against the Physics Unit Standards this year?

   Yes No

   b) If you assessed against the Physics Unit Standards last year but are not this year, please state your main reason for discontinuing.

____

____

2. A valid course result is one that accurately describes student achievement relative to the course objectives.

a) How valid do you think a Sixth Form Certificate Grade is for indicating student achievement in Sixth Form Physics?

Very valid	Valid	Not sure	Invalid	Very invalid
5	4	3	2	1

Comment:

b) How valid do you think Unit Standard credits are for indicating student achievement in Sixth Form Physics?

Very valid	Valid	Not sure	Invalid	Very invalid
5	4	3	2	1

Comment:

3) Which of the following best describes your attitude towards the use of Unit Standards in Sixth Form Physics?

Very enthusiastic	Enthusiastic	Not sure	Unenthusiastic	Very unenthusiastic
5	4	3	2	1

Comment:

4) Please indicate to what extent you agree or disagree with the following statements:

a) Unit standards are appropriate for assessing practical work skills.

Strongly agree	Agree	Not Sure	Disagree	Strongly disagree
5	4	3	2	1

Comment:____

b) Unit standards are appropriate for assessing students' abilty to solve physics problems.

Strongly agree	Agree	Not Sure	Disagree	Strongly disagree
5	4	3	2	1

Comment:____

c) Unit standards are appropriate for assessing students' ability to explain physics concepts.

Strongly agree	Agree	Not Sure	Disagree	Strongly disagree
5	4	3	2	1

Comment:____

d) Do you consider the split between competent, not yet competent for each performance criterion to be sufficient?

Yes No

e) If your answer to d) was No, how many achievement levels would you prefer?

5 4 3

f) Criteria for excellence should be built into each Unit Standard.

Strongly agree	Agree	Not Sure	Disagree	Strongly disagree
5	4	3	2	1

Comment:____

____

____

____

Curriculum coverage

5. The following list of topics is based on the curriculum level 7 content described in *Physics in the New Zealand Curriculum*. Please place a tick beside each topic you will have completed with your class by the end of this year.

Topic:	
Kinematic equations	
Relative motion	
Vectors	
Projectile motion	
Circular motion	
Levers and torque	
Momentum	
Impulse	
Elastic and inelastic collisions	
Hooke's law	
Elastic potential energy	
Parallel and series circuits	
Ohms law	
Voltage dividers	
Non-ohmic conductors	
Magnetic force on a current carrying conductor	
Magnetic force on a charged particle	
Electromagnetic induction	
Simple generator	
Reflection and refraction of light waves	
Reflection and refraction of water waves	

Superposition of waves	
The electromagnetic spectrum	
Atomic models	
Half life	
Radioactive decay	
Alpha, beta and gamma emission	

Assessment

6. For your sixth form physics class, what is the total number of hours of class time you will spent this year on assessment? Include exams, tests, assessed practical work etc. If not applicable, write NA in the box.

a) Sixth Form Certificate hours

b) Unit Standards hours

7. Estimate the average amount of time (in minutes) you spend each week on the following tasks associated with the assessment and administration of sixth form assessment. You may add extra tasks if necessary. If not applicable, write NA in the box.

Task	Sixth Form Certificate Time (minutes)	Unit Standards Time (minutes)
Writing (re)assessment activities		
Marking		
Moderation		
Department meetings		
Other		

8. What procedures are used within your physics department for ensuring that the assessment activities used for different Year 12 physics classes are of a comparable standard?

9. What procedures are used within your physics department for ensuring that the assessor judgements of teachers of different Year 12 physics classes are consistent?

____

____

____

10. (a) Please rank the following options for physics assessment at the sixth form level in order of preference. Assign a 1 to your most preferred option and a 5 to you least preferred option.

Assessment option	**Rank**
National examination only	
Sixth Form Certificate (present model)	
Sixth Form Certificate (moderated using reference tests)	
Unit Standards only	
Sixth Form Certificate (present model) for problem solving and Unit Standards for practical work.	

Other (Please state):____

(b) Please rank the following options for physics assessment at the seventh form level in order of preference. Assign a 1 to your most preferred option and a 4 to you least preferred option.

Assessment option	**Rank**
National Bursary / Scholarship examination only	
National Bursary / Scholarship examination with an internally assessed component scaled to the exam.	
Unit Standards only	
A combination of examination and Unit Standards both of which contribute credit towards the National Certificate.	

Other (Please state):____

Section B

To be answered only by teachers who are currently assessing against the Physics Unit Standards

11. a) Please circle the level(s) at which the physics department in your school is currently assessing against the Physics Unit Standards.

Level 1 Level 2 Level 3

b) Please indicate the number of physics unit standard credits you will offer in your assessment programme at levels 2 and 3 in 1998.

Level	Number of credits
2	
3	

12. How well do you think the level 2 Physics Unit Standards reflect the Physics Curriculum?

Very well	Quite well	Not sure	Poorly	Very poorly
5	4	3	2	1

Comment:____

13. How useful do you find the performance criteria within each element for indicating student mastery of the learning outcome stated in the element?

Very useful	Useful	Not sure	Of limited use	No use
5	4	3	2	1

Comment:____

14. How do you ensure that the project and practical work handed in by students for assessment against the physics unit standards is authentic?

____

____

____

15. How do you think your students feel about Unit Standards assessment in physics?

Very positive	Positive	Not sure	Negative	Very negative
5	4	3	2	1

Comment:____

____

____

____

16. How do you think assessing against the Physics Unit Standards has affected student motivation in your physics course?

Very favourably	Favourably	Not sure	Unfavourably	Very unfavourably
5	4	3	2	1

Comment:____

____

____

____

Moderation

17. How satisfactory do you think each of the following aspects of the physics moderation system is in achieving comparability between schools?

a) Moderation of assessment activities

Very satisfactory	Satisfactory	Not sure	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:____

____

____

____

b) Moderation of assessor judgements

Very satisfactory	Satisfactory	Not sure	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:____

____

____

____

c) The pre-moderated assessment activities provided by NZQA this year

Very satisfactory	Satisfactory	Not sure	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:____

____

____

____

d) Communication with your local moderator

Very satisfactory	Satisfactory	Not sure	Unsatisfactory	Very unsatisfactory
5	4	3	2	1

Comment:____

____

____

____

18. How useful do you find the assessment activities in the *Assessment Guide: Physics*?

Very useful	Useful	Not sure	Of limited use	No use
5	4	3	2	1

Comment:____

19. Which of the following best describes the effect assessing against the Physics Unit Standards has had on your classroom teaching?

Very positive	Positive	Not sure	Negative	Very negative
5	4	3	2	1

Comment:____

20. Which of the following best describes the effect of assessing against the Physics Unit Standards on student learning:

a) for the more able student?

Very positive	Positive	Not sure	Negative	Very negative
5	4	3	2	1

b) for the average student?

Very positive	Positive	Not sure	Negative	Very negative
5	4	3	2	1

c) for the less able student?

Very positive	Positive	Not sure	Negative	Very negative
5	4	3	2	1

21. Do you think assessment against the Physics Unit standards enables recognition of excellence?

Yes No

Comment:____

22. Please state any suggestions you may have about how criteria for recognising excellence can be built into assessment against Unit Standards.

23. How helpful was the communication you received from NZQA in explaining the administrative procedures associated with assessment and moderation of the Physics Unit Standards?

Very useful	Useful	Not sure	Of limited use	No use
5	4	3	2	1

Comment:____

24. What is your preferred method for receiving a copy of the Unit Standards form NZQA ?

The c.d. Framework Explorer A paper copy

Comment:____

25.a) Do you agree with the decision not to have a Nationally PrescribedActivity in 1998?

Yes No

Comment

b) Do you support the provision of the pre-moderated assessment activities this year?

26. If there are any skills or content areas in the physics curriculum which in your opinion, cannot be expressed as learning outcomes and stated as an element please list these below.

27. Do you have any additional comments about assessment against Unit Standards, assessment in sixth form physics or implementation of the National Qualifications Framework? (Continue on the back of this page if necessary)

Comment:

Thank you for taking part in this survey. Please return the completed questionnaire in the postage paid envelope provided to

John Boereboom
Christchurch College of Education
PO Box 31 065
Christchurch.

Appendix 16

Level 2 Nationally Prescribed Activity and sample student answer for 1996 end-point assessor judgement agreement trial

New Zealand Qualifications Authority

NATIONAL QUALIFICATIONS FRAMEWORK
MODERATION OF PHYSICS UNIT STANDARDS, 1996

NATIONALLY PRESCRIBED ACTIVITY FOR PHYSICS UNIT STANDARDS

LEVEL 2

QUESTION AND ANSWER BOOKLET

Time allowed: 1 hour

This assessment activity assesses:

Unit: 6380 Apply formulae, graphical and vectorial methods to find unknowns for a physical system.

Element:
1. Recognise the principle(s) and quantities of a physical system.
2. Determine an unknown quantity using a formula.
3. Determine an unknown quantity using a graph.
4. Determine an unknown quantity using two dimensional vectors.
   Range: vector addition, vector subtraction.

INSTRUCTIONS

Check that this booklet contains pages 2 – 7 in the correct order.
Answer all questions.
Complete the tasks in the spaces provided in this booklet. Show all relevant working.
Numerical answers must be rounded to an appropriate number of significant figures and include correct units.
Answers to questions must be stated in the context of the problem using a complete sentence.

Assessor

After assessing the student's answers please indicate in the boxes provided on the right if the standard for each element has been achieved or not.

Element:			
	1.	Recognise the principle(s) and quantities of a physical system.	☐
	2.	Determine an unknown quantity using a formula.	☐
	3.	Determine an unknown quantity using a graph.	☐
	4.	Determine an unknown quantity using two-dimensional vectors (vector addition, vector subtraction).	☐

AT THE END OF THE TIME FOR COMPLETING THIS ACTIVITY, HAND THIS BOOKLET TO THE SUPERVISOR

(Turn over

PHYSICS IN ANTARCTICA

QUESTION ONE

Scientists working at Scott Base often use a snowmobile to visit research sites.

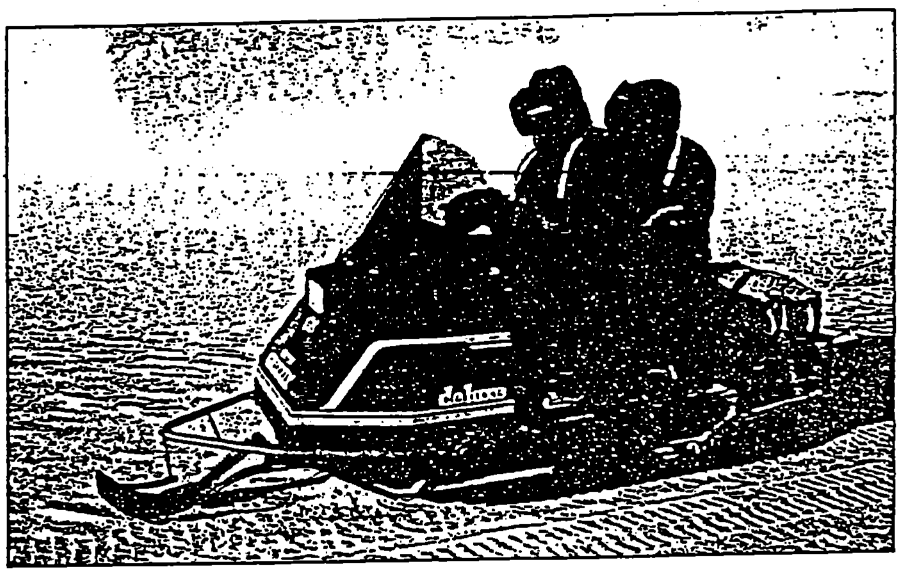
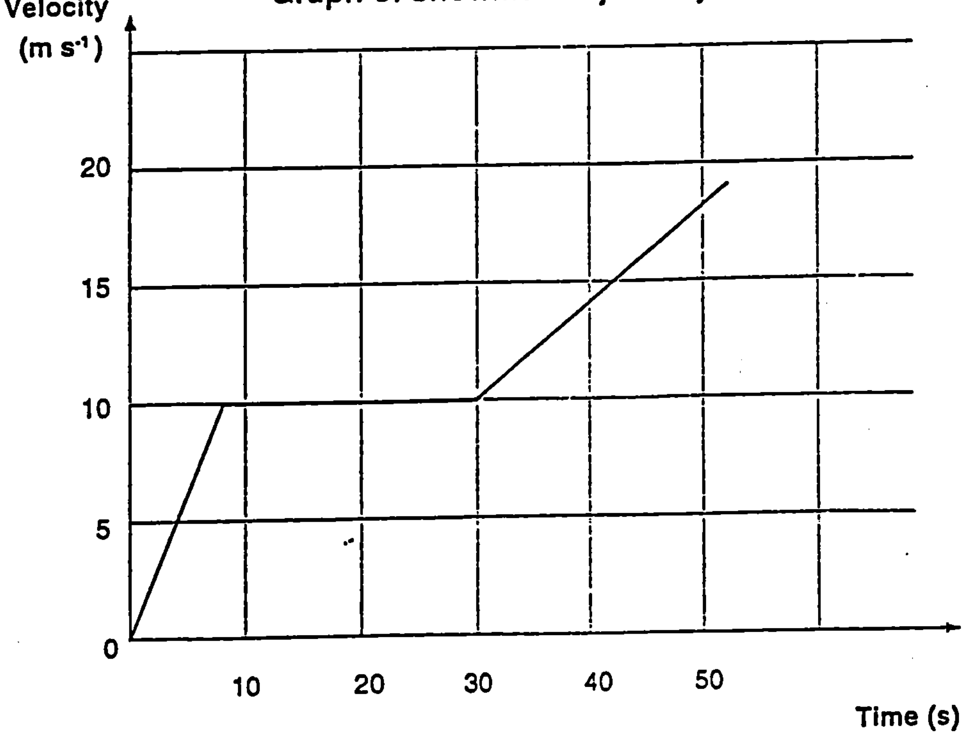


Fig 1: Snowmobile at Scott Base

This is the velocity versus time graph for part of the journey of a snowmobile of mass 1650 kg. The snowmobile starts from rest.

Graph of snowmobile journey



(a) What is the speed of the snowmobile 42 seconds after starting from rest?

The speed of the snowmobile is $15.0\ m\,s^{-1}$.

☐ 3.1 ☐ 3.2

(b) What is the acceleration of the snowmobile 34 seconds after the start of the journey?

$a = \text{slope} = \frac{18-10}{50-30} = \frac{8}{20} = 0.4\ m\,s^{-1}$

☐ 3.1 ☐ 3.2

(c) What is the total distance travelled by the snowmobile after 52 seconds?

$d =$ area under the graph

$= \triangle + \square + \triangle$

$= \frac{1}{2} \times 8 \times 10 + 42 \times 10 + \frac{1}{2} \times 22 \times 9$

$= 40 + 420 + 99$

$= 559\ m$

☐ 3.1 ☐ 3.2

(d) Calculate the average velocity of the snowmobile for the journey shown on the graph.

$v =$ total distance / total time

$= 559 / 52$

$= 10.8$

The average velocity is $11.0\ m\,s^{-1}$

☐ 2.1 ☐ 2.2 ☐ 2.3 ☐ 2.4 ☐ 2.5

(e) Calculate the increase in the kinetic energy of the snowmobile when it accelerates from 12 to 18 m s⁻¹.

$E_k = \frac{1}{2} m (v_f^2 - v_i^2)$

$E_k = \frac{1}{2} \times 1650 \times (18^2 - 12^2)$

$E_k = 148\,500$

The increase in kinetic energy is 148000 J

☐ 2.1 ☐ 2.2 ☐ 2.3 ☐ 2.4 ☐ 2.5

(Turn over

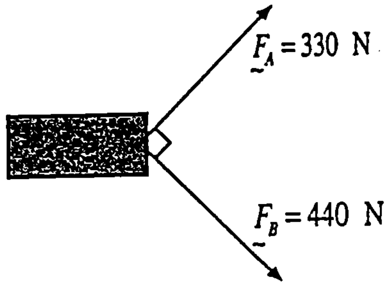
QUESTION TWO

Before snowmobiles were invented, Antarctic explorers used husky dogs to pull their sledges.

Fig 2: Dog team in action

Two huskies are pulling a sledge with forces of 330 N and 440 N as shown in the diagram.

The ropes from the dogs to the sledge are at right angles to each other.



(a) State a vector equation which can be used to calculate the resultant force exerted by the huskies on the sledge. Identify any symbols you use.

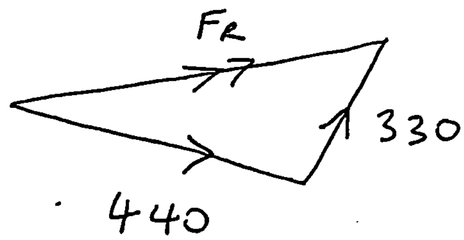
$\underset{\sim}{F}_A$ = husky A force

$\bar{F}_R = \bar{F}_A + \bar{F}_B$ $\quad$ $\underset{\sim}{F}_B$ = husky B force

$\underset{\sim}{F}_R$ = total force

☐ 4.1 ☐ 4.2

(b) Draw a labelled vector diagram which shows the addition of the forces exerted by the huskies on the sledge and the resultant force.



☐ 4.3

(c) Calculate the magnitude of the resultant force and the angle it makes with the rope in which the force is 440 N.

State your answer in the form of a sentence.

$F_R = \sqrt{440^2 + 330^2}$

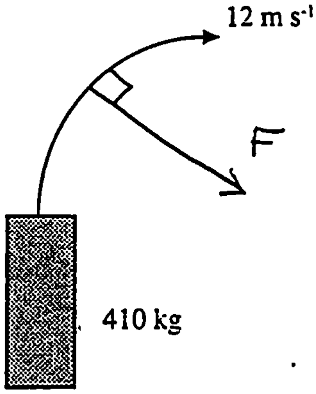
$F_R = 550$ N

$\theta = \tan^{-1} \frac{330}{440}$

$\theta = 37°$

4.4

(d) A sledge of mass 410 kg travels around a circular bend of radius 10 m with a speed of 12 m s^{-1}. On the diagram below, draw a labelled vector that represents the resultant force which acts on the sledge.



1.1

(e) The formula which can be used to calculate this force is

$$F = \frac{mv^2}{R}$$

For each symbol used in the equation state what it represents, its value in the problem and the S.I. unit it is measured in. Record your answers in the table below.

Symbol	What the symbol represents	Value	S.I. unit
F	centripetal force	5900	N
m	mass	410	kg
v	velocity	12	m s^{-2}
R	radius of bend	10	m

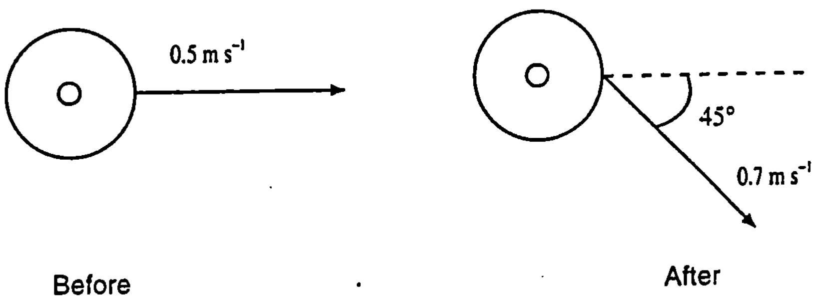
1.2

(Turn over

QUESTION THREE

To pass the time during the Antarctic winter, the team of scientists at Scott Base play a game of curling. Curling is a form of bowls on ice in which a curling stone is pushed across the ice.

A curling stone of mass 18 kg moves across the ice with a velocity of 0.5 m s^{-1} in a direction due East. The stone hits an obstacle embedded in the ice and moves off with a velocity of 0.7 m s^{-1} ($0.5\sqrt{2} \text{ m s}^{-1}$) in a direction 45° South of East as shown in the diagram below.



(a) Calculate the momentum of the curling stone before it hits the obstacle.

$\underset{\sim}{P}_{before}$ = 18 × 0.5

= 9

4.1

(b) Calculate the momentum of the curling stone after it hits the obstacle.

$\underset{\sim}{P}_{after}$ = 18 × 0.707

= 12.7 kg m s^{-1} 45° S of E

4.1

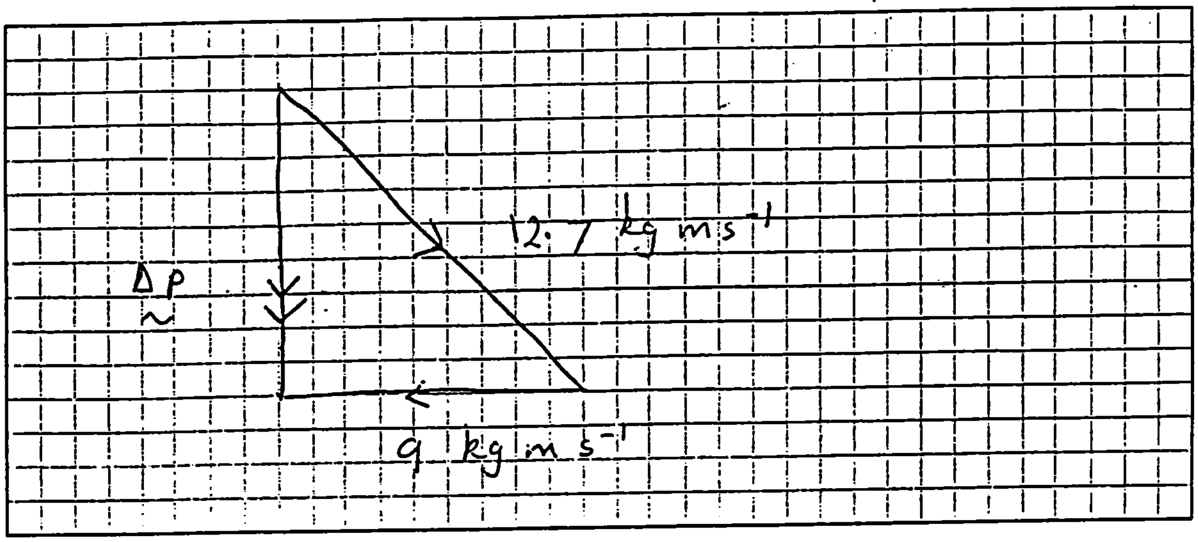
(c) State a vector equation from which the change in momentum can be calculated.

$\Delta \underset{\sim}{P}$ = 12.7 kg.ms^{-1} − 9 kgms^{-1}

= 3.7 kg m s^{-1}

4.2

(d) Draw a labelled vector diagram which shows the change in momentum of the curling stone.



4.3

(e) Calculate the change in momentum of the curling stone.

$\Delta p = \sqrt{12.7^2 - 9^2} = 9$ kg m s^{-1} South

4.4

NATIONALLY PRESCRIBED ACTIVITY FOR PHYSICS UNIT STANDARDS LEVEL 2, 1996
UNIT STANDARD 6380
ASSESSMENT SCHEDULE

Assessment schedule : Question 1

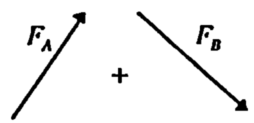
Task Number	Performance Criteria	Evidence	Judgement
a	3.1 3.2	The speed of the snowmobile 42 s after the start of the journey is 15 m s^{-1}.	Allow answer in the range 14.5 - 15.5 m s^{-1}. (3.1) Must have the correct unit. (3.2)
b	3.1 3.2	acceleration = slope of the graph at $t = 34\ s$ $a = \frac{18-10}{50-30}$ $a = 0.4 \text{ m s}^{-2}$	Answer must be in range 0.4 – 0.42. (3.1) The answer must link the acceleration to the slope of the graph and be accompanied by the correct unit. (3.2)
c	3.1 3.2	The total distance travelled is equal to the area under the graph. $d = \frac{1}{2} \times 8 \times 10 + 44 \times 10 + \frac{1}{2} \times 22 \times 9$ $d = 40 + 440 + 99$ $d = 579$ m	The answer must link the total distance travelled to the area under the graph. (3.2) The area under the graph must be calculated correctly. (3.1) The answer and unit must be correct. (3.2)
d	2.1 2.2 2.3 2.4 2.5	average velocity = total distance travelled / total time taken $= \frac{579}{52}$ $= 11.1 \text{ m s}^{-1}$ The average velocity of the snowmobile is 11 m s^{-1}.	A formula from which the average velocity can be calculated must be stated. (2.1) Correct working must be shown. (2.2) The answer must be correct. (2.3) OR Answer to c divided by 52. The answer must be stated in the form of a sentence using correct units. (2.4) Answer must be rounded to two significant figures. (2.5)

Assessment schedule : Question 1 — Continued

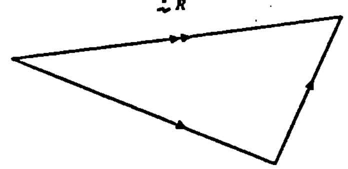
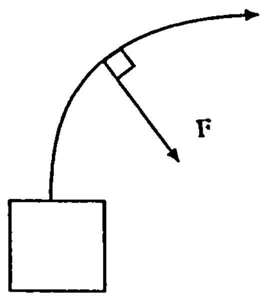
c	2.1	$E_k = \frac{1}{2}m\left(v^2{}_f - v^2{}_i\right)$	A formula from which the increase in kinetic energy can be calculated must be stated. (2.1)
	2.2	$E_k = \frac{1}{2} \times 1650 \times \left(18^2 - 12^2\right)$	Correct working must be shown. (2.2)
	2.3	$E_k = 148500$ J	The answer must be correct. (2.3)
	2.4		Answer must be stated in the form of a sentence using correct units. (2.4)
	2.5	The increase in the kinetic energy of the snowmobile is 150 000 J	Answer must be rounded to two significant figures. (2.5)

Explanatory notes:
1. Performance criteria 3.1 and 3.2 must be met in task a (interpolation) ***and*** task b (gradient) ***and*** task c (area) for credit of element 3
2. Performance criteria 2.1-2.5 may be met in ***either*** task d ***or*** task c for credit of element 2.

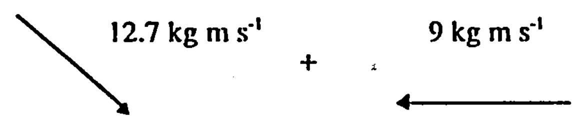
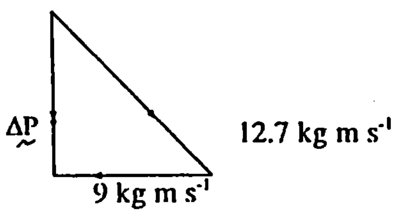
Assessment schedule: Question 2

Task Number	Performance criteria	Evidence	Judgement
a	4.1	$\underset{\sim}{F}_A$ is the force exerted by husky A on the sledge. $\underset{\sim}{F}_B$ is the force exerted by husky B on the sledge.	All vector quantities must be identified correctly. (4.1)
	4.2	$\underset{\sim}{F}_R$ is the resultant force. $\underset{\sim}{F}_A + \underset{\sim}{F}_B = \underset{\sim}{F}_R$ OR $\underset{\sim}{F}_R =$ 	A vector equation is stated from which the resultant can be calculated. (4.2) The equation may be either in symbolic or vector format

Assessment schedule: Question 2 — Continued

b	4.3	$\underset{\sim}{F}_R$  $\underset{\sim}{F}_B = 440$ N $\underset{\sim}{F}_A = 330$ N	All vectors must be labelled with symbol and or quantity. The vectors must be added head to tail. A sketch diagram is sufficient but the 440 N vector must be longer. The arrowhead on the resultant must point in the right direction.
c	4.4	$\underset{\sim}{F}_R = \sqrt{440^2 + 330^2}$ $\underset{\sim}{F}_R = 550N$ $\theta = \tan^{-1}\dfrac{330}{440}$ $\theta = 36.9°$ or $37°$ or 0.65 rad.	The magnitude and unit of $\underset{\sim}{F}_R$ must be correct. The value for θ must be correct. (Allow answer in radians or degrees).
d	1.1		The labelled force must act at right angles to the path of the sledge and point towards the centre of the circular arc. (1.1)
e	1.2	Symbol / What it represents / Value / Unit F / centripetal force / 5900 or 5904 / N m / mass of the sledge / 410 / kg v / velocity of the sledge / 12 / m s^{-1} R / radius of the bend / 10 / m	All symbols must be correctly identified with the correct value and unit. (1.2) (Do not penalise wrong no. of significant figures)

Assessment schedule: Question 3

Task Number	Performance criteria	Evidence	Judgement
a	4.1	$p_{before} = 18 \times 0.5$ $= 9$ kg m s^{-1} East	Must have the correct value and direction
b	4.1	$p_{after} = 18 \times 0.707$ $= 12.7$ kg m s^{-1} 45° South of East (13 kg m s^{-1})	Must have the correct value and direction Allow answer in range 12.6 – 13.0 kg m s^{-1}
c	4.2	$\Delta \underset{\sim}{P} = \underset{\sim}{P}_{after} - \underset{\sim}{P}_{before}$ $\Delta \underset{\sim}{P} = \underset{\sim}{P}_{after} + (-\underset{\sim}{P}_{before})$ or 	The equation may be in either symbolic or vector format. The equation must convey the idea that subtracting a vector is equivalent to adding the vector in the opposite direction.
d	4.3		All vectors must be labelled with symbol and/or quantity. Arrowheads on the vectors must point in the right direction. (Units optional)
e	4.4	The change in momentum of the curling stones is, $\Delta \underset{\sim}{P} = \sqrt{12.7^2 - 9^2} = 9$ kg m s^{-1} South= 9 kg m s^{-1}	Allow answers in the range 8.8 - 9.1 kg m s^{-1} Answer must include the correct SI unit and direction.

Results of the 1996 end-point assessor judgement agreement trial

Task	Performance criterion assessed	NM decision	Number of assessors who judged the student answer to meet the performance criterion	Number of assessors who judged the student answer not to meet the performance criterion	Percentage of end-point assessor judgements which agreed with the consensus decision
1(a)	3.1	Yes	43	0	100
	3.2	Yes	43	0	100
1(b)	3.1	Yes	43	0	100
	3.2	No	5	38	88
1(c)	3.1	No	12	31	72
	3.2	No	18	25	58
1(d)	2.1	Yes	43	0	100
	2.2	Yes	43	0	100
	2.3	Yes	42	1	98
	2.4	Yes	43	0	100
	2.5	No	8	35	81
1(e)	2.1	Yes	43	0	100
	2.2	Yes	43	0	100
	2.3	Yes	43	0	100
	2.4	Yes	43	0	100
	2.5	No	2	41	95
2(a)	4.1	Yes	40	3	93
	4.2	Yes	43	0	100
2(b)	4.3	Yes	33	10	76

2(c)	4.4	Yes	42	1	98
2(d)	1.1	Yes	43	0	100
2(e)	1.2	No	7	36	84
3(a)	4.1	No	3	40	93
3(b)	4.1	Yes	41	2	95
3(c)	4.2	No	2	41	95
3(d)	4.3	Yes	43	0	100
3(e)	4.4	Yes	43	0	100
Mean percentage of agreement of end-point assessor judgements					94
Standard deviation of percentages of agreement of end-point assessor judgements					11

Appendix 17

Level 2 Nationally Prescribed Activity and sample student answer for 1997 end-point assessor judgement agreement trial

New Zealand Qualifications Authority

NATIONAL QUALIFICATIONS FRAMEWORK
MODERATION OF PHYSICS UNIT STANDARDS, 1997

NATIONALLY PRESCRIBED ACTIVITY FOR PHYSICS UNIT STANDARDS LEVEL 2

QUESTION AND ANSWER BOOKLET

Time allowed: 45 minutes

This assessment activity assesses:

Unit Standard 6382 **Demonstrate knowledge of waves.**

Element: 1. Describe waves.
Range: waves on springs and strings, water waves.
2. Determine an unknown quantity when a wave refracts.
Range: two of - wave velocity, frequency, wavelength, period.

USEFUL FORMULAE: $n_1v_1 = n_2v_2$ $v = f\lambda$

INSTRUCTIONS:

Print your name, your Physics teacher's name and the name of your provider/school in the spaces provided above.
Check that this booklet contains pages 2 - 10 in the correct order.
Answer all tasks and complete them in the spaces provided in this booklet. Show all relevant working.
Numerical answers must include correct SI units and the appropriate number of significant figures.

Assessor

After assessing the student's answers please indicate in the boxes provided on the right if the standard for each element has been achieved or not.

Element:			✓/✗
	1.	Describe waves.	
	2.	Determine an unknown quantity when a wave refracts.	

✓ = the standard for the element has been achieved.
✗ = the standard for the element has not been achieved.

AT THE END OF THE TIME FOR COMPLETING THE ACTIVITY, HAND THIS BOOKLET TO THE SUPERVISOR

(Turn over

TASK ONE: WATER WAVES

Emma, a physics teacher, is talking to her 15 year-old brother Jamie, who has just come in from the beach near their home.

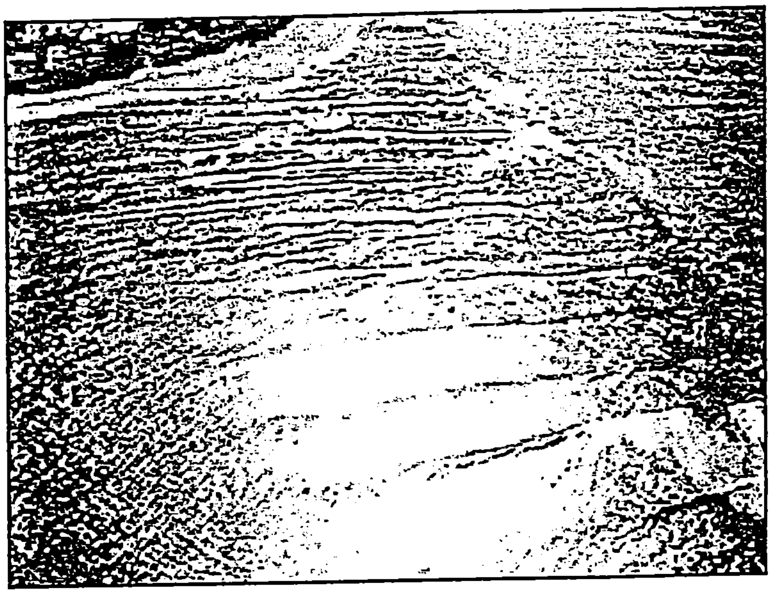


Fig. 1

Jamie: "I was standing on the cliff above the beach watching the waves coming in from far away out at sea. As they got closer to the shore, they seemed to bend."

Emma: "It's called refraction - the waves change direction when they travel from deep to shallow water. You've done this in physics at school haven't you? Did you notice any other changes to the waves as they got closer to the beach?"

Emma sketched the diagram below showing wavefronts incident on the deep / shallow water boundary.

Waves incident at a boundary between deep and shallow water

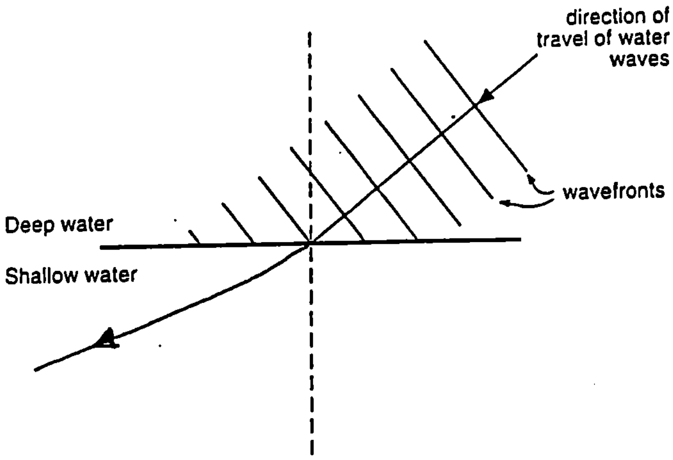


Fig. 2

Emma: "What happened to the speed, wavelength, direction and frequency as they got closer to the beach?

Jamie tried hard to think back to what he did in class about water waves. Can you help him?

(a) 1. On Fig. 2 (page 2) draw an arrow to show the direction of travel of the water waves in the shallow water. ☐ 1.1

State how each of the following quantities is affected as the waves enter the shallow water.

2. Speed decreases ☐ 1.1

3. Wavelength increases. ☐ 1.1

4. Frequency remains unchanged. ☐ 1.1

Jamie: "I also noticed that some of the waves were reflected from that long flat cliff face at the end of the beach. It was just like the reflections I saw when we used the ripple tank at school."

Emma: "Can you draw what happened to show how the waves were reflected?"

Emma sketched the diagram below showing wavefronts incident on a flat barrier like the cliff face. Help Jamie to answer his sister's question.

(b) On Fig. 3 (below)

1. Draw an arrow to show the direction of travel of the water waves after reflection from this barrier. ☐ 1.1

2. Label (i) the angle of incidence, and (ii) the angle of reflection. ☐ 1.1

3. Draw at least four of the reflected wavefronts. ☐ 1.1

Reflection of water waves from a flat barrier

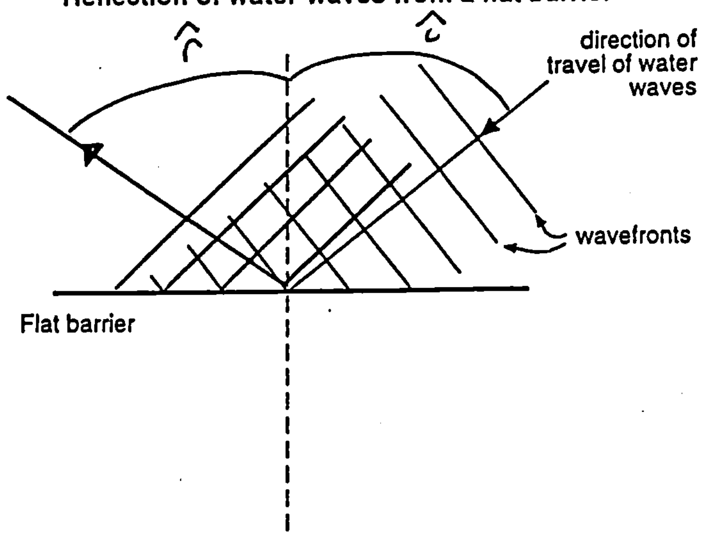


Fig. 3

(Turn over

Emma had taken some measurements on that beach for her own lessons on water waves. She had noted that twelve waves reached the beach in four minutes and that the crests of the incoming waves were 5.0 m apart.

Using Emma's information

(c) 1. Determine the frequency of the waves in the shallow water. In the answer state the correct SI unit.

$f = \frac{12 \text{ waves}}{4 \text{ minutes}} = \frac{12}{4 \times 60} = 0.050 \text{ Hz}$

Frequency = 0.050 Hz

☐ 2.1 ☐ 2.2

2. Determine the speed of the waves in the shallow water. In the answer state the correct SI unit.

$v = f \times \lambda$

$= 0.50 \times 5.0$

Speed = 0.25 m s^{-1}

☐ 2.1 ☐ 2.2

TASK TWO: STRING WAVES

Michelle and Sam are playing 'snakes' on the floor with a long rope.

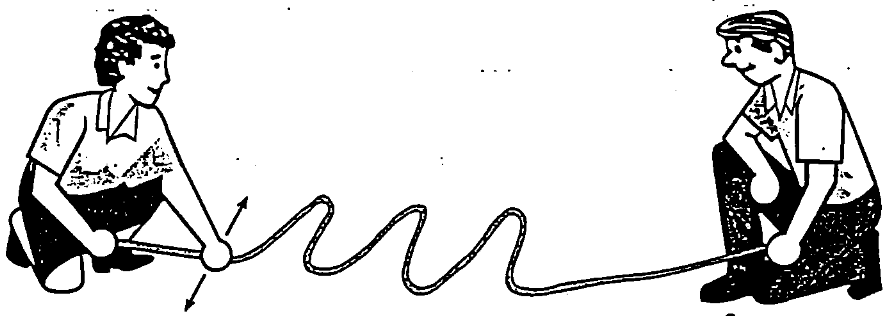


Fig. 4

(a) Sam holds one end tightly and Michelle sends a single wave pulse along the rope towards him (see Fig. 5(a)).

Wave pulse sent along the rope to a fixed end

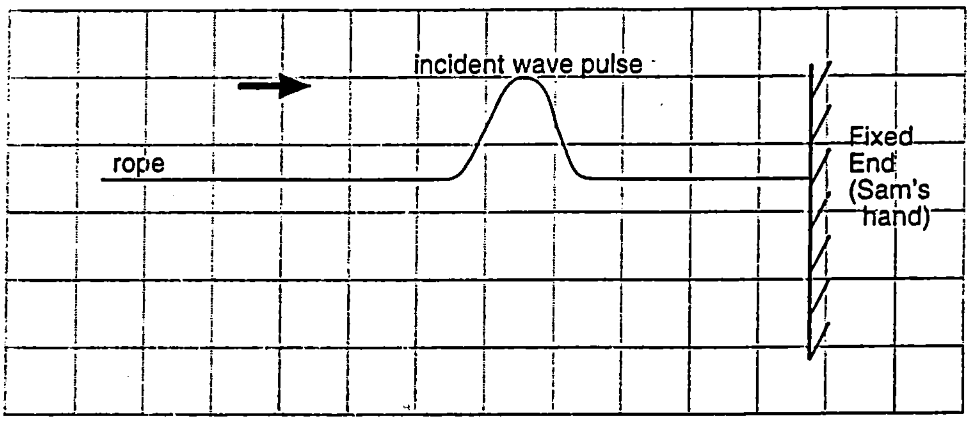


Fig. 5(a)

When the pulse reaches Sam it is reflected from his hand.

1. On Fig. 5(b), below sketch the reflected wave pulse. Assume the amplitude of the reflected wave pulse is the same as the amplitude of the incident wave pulse.

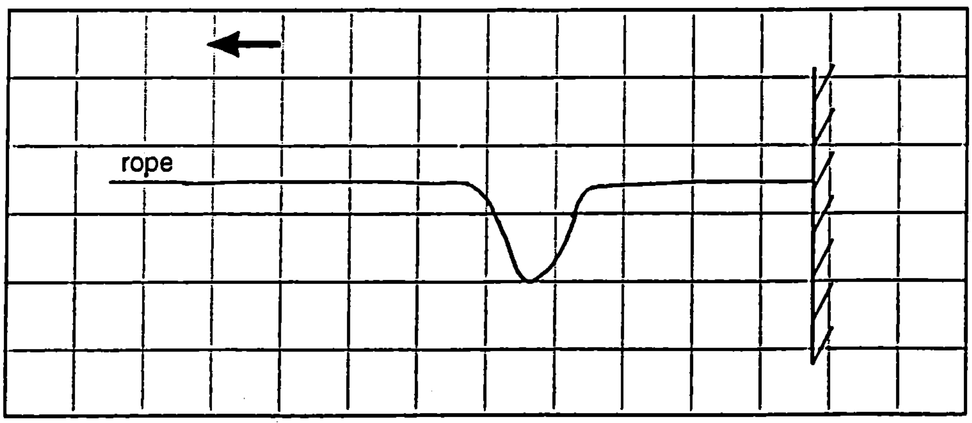


Fig. 5(b)

☐ 1.1

(Turn over

2. Usually, the amplitude of the reflected wave pulse would be less than the amplitude of the incident wave pulse. Explain why this would happen.

because of the energy loss on reflection.

☐ 1.1

(b) Sam and Michelle now both send single wave pulses along the rope to each other.

1. Fig. 6(a) shows two pulses of equal amplitude 2.0 cm, approaching each other.

Wave pulses sent in opposite directions along the rope

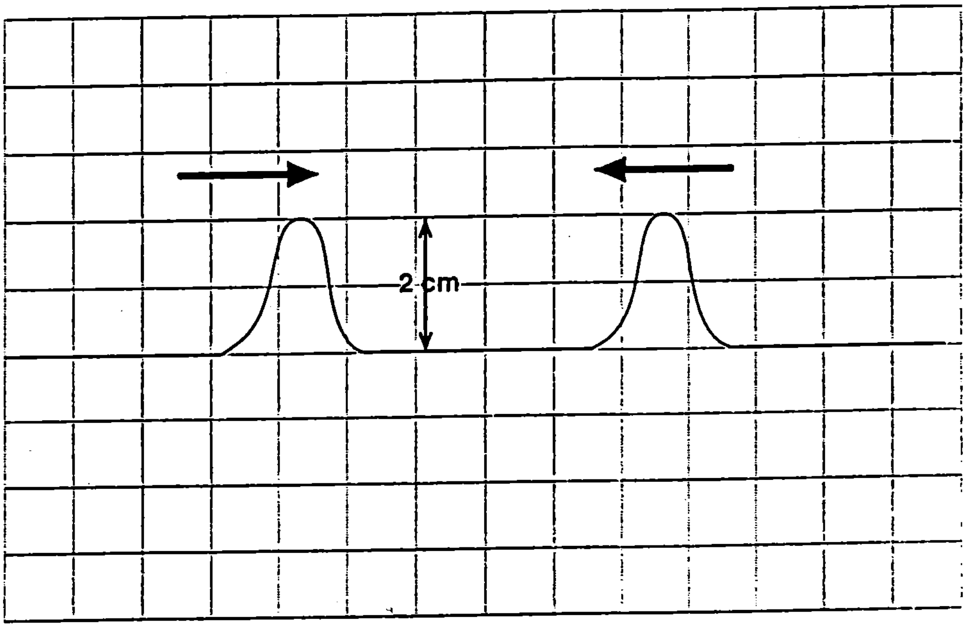


Fig. 6(a)

On Fig. 6(b) below sketch the resulting waveform when the centres of the two pulses overlap.

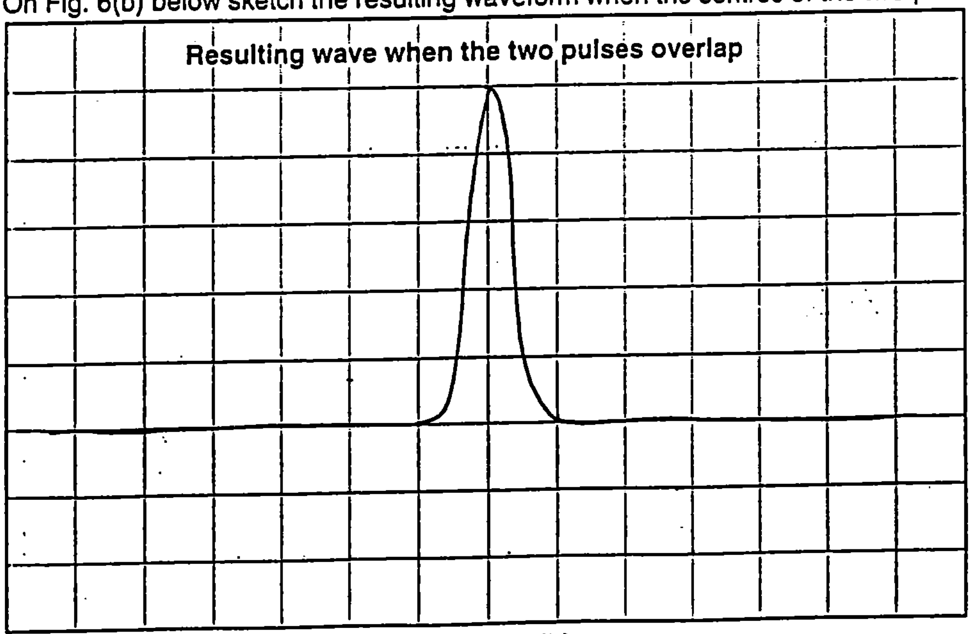


Fig. 6(b)

☐ 1.2

2. Fig. 7(a) shows one pulse of amplitude 3.0 cm and a second pulse of amplitude 2.0 cm approaching each other on opposite sides of the rope.

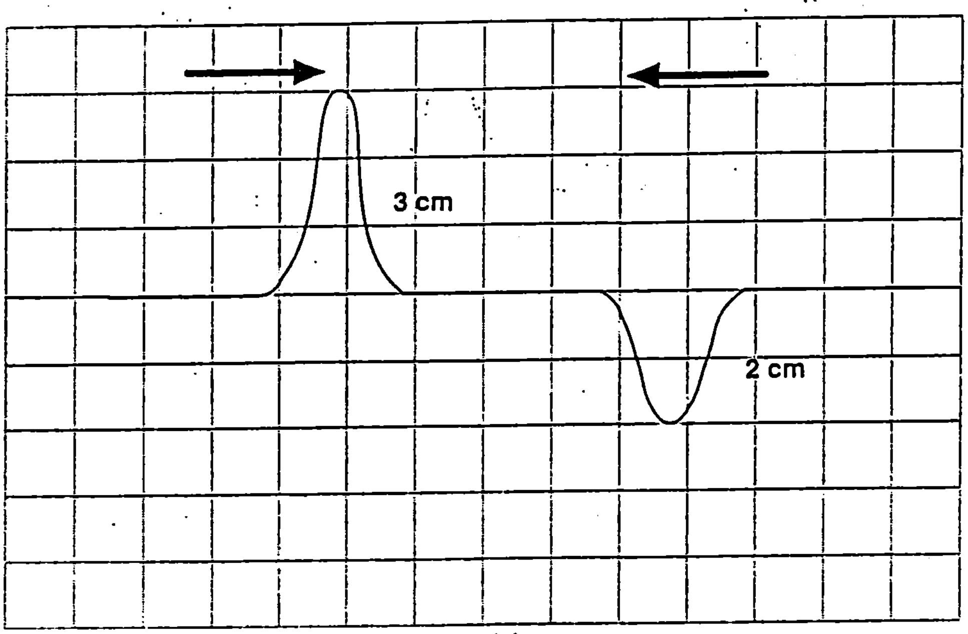


Fig. 7(a)

On Fig. 7(b) below sketch the resulting waveform when the centres of the two pulses pass the same point on the rope.

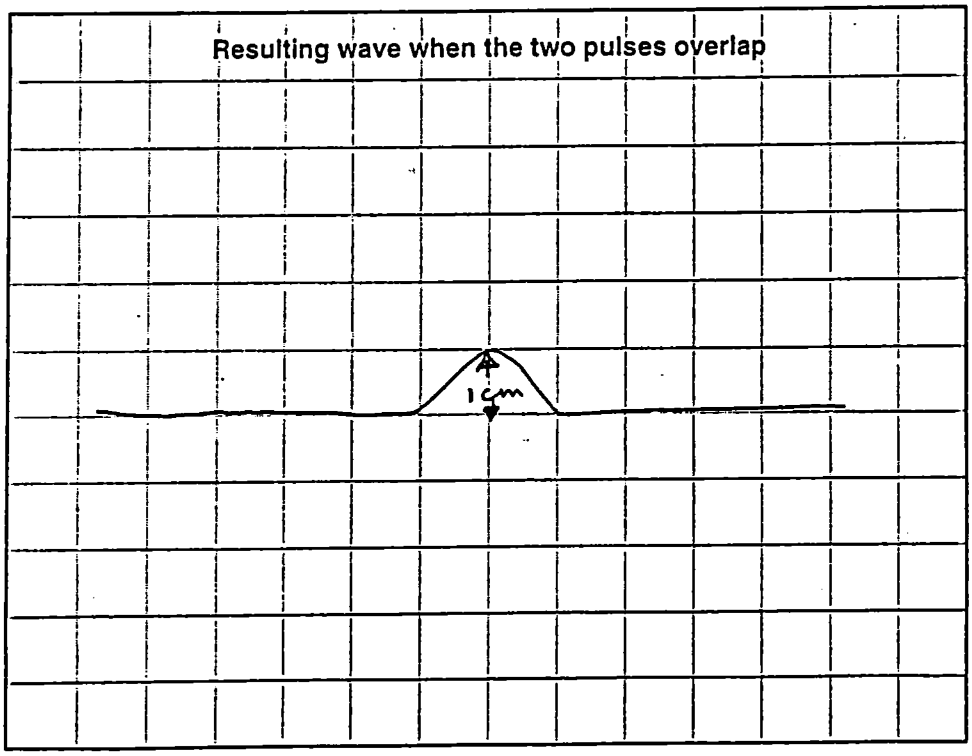


Fig. 7(b)

☐ 1.2

(c) Michelle now ties a lighter piece of string to the rope. Sam holds the string and Michelle holds the heavier rope. Michelle again sends a single wave pulse along the rope towards Sam. This incident wave pulse arrives at the junction between the rope and the string as shown in Fig. 8 below.

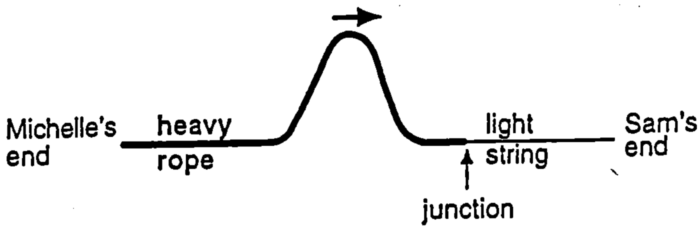


Fig. 8

At the junction the incident wave pulse produces a reflected pulse and a refracted pulse.

Key Word List

Amplitude	**Phase Change**	**Direction**

1. Describe the appearance of the **refracted pulse** compared with the incident wave pulse, using the terms from the key word list above.

No phase change, smaller amplitude travelling to the right.

☐ 1.1

2. Describe the appearance of the **reflected pulse** compared with the incident wave pulse, using the terms from the key word list above.

Smaller amplitude, 180° phase change travelling to the left.

☐ 1.1

TASK THREE: RAINBOWS AND REFRACTION

A rainbow is often seen after a shower of rain. (Fig. 9).
The various colours in the rainbow can be seen because light of different frequencies travels through raindrops at different speeds. This causes each colour to follow a separate path as shown in Fig. 10. This is a simple model of how a rainbow is formed.

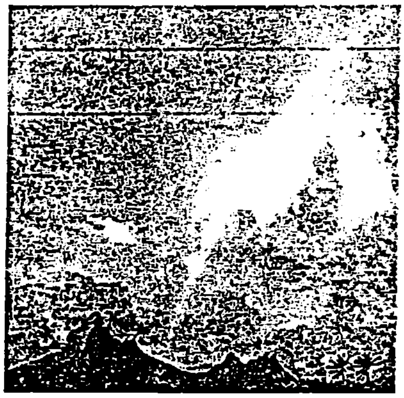


Fig. 9

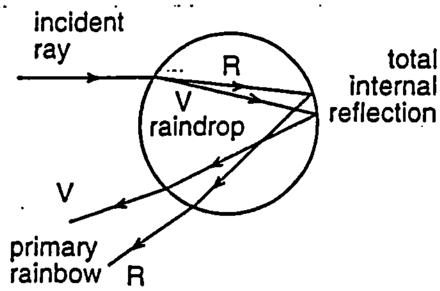


Fig. 10

Consider the passage of a ray of light through the raindrop as shown in Fig.10.

Refractive index of air	= 1.00
Refractive index of water for red light	= 1.33
Refractive index of water for blue light	= 1.34
Speed of light in air	= 3.00×10^8 m s^{-1}

(a) Using the information given above, determine the following:

1. The speed of red light through the raindrop. State your answer to three significant figures with the appropriate SI unit.

$n_1 v_1 = n_2 v_2$

$1 \times 3 \times 10^8 = 1.33 \times v_2$

$v_2 = 3 \times 10^8 / 1.33$

Speed of red light = 2.3×10^8 m s^{-1}

☐2.1 ☐2.2

2. The speed of blue light through the raindrop. State your answer to three significant figures giving the appropriate SI unit.

$v_2 = 3 \times 10^8 / 1.34$

Speed of blue light = 2.2×10^8 m s^{-1}

☐2.1 ☐2.2

(Turn over

(b) Determine the frequency of the red light which has a wavelength of 6.00×10^{-7} m in air. In the answer state the correct SI unit.

$f = \frac{\lambda}{v}$

$= 6 \times 10^{-7} / 3 \times 10^{8}$

$= 2 \times 10^{-15}$ Hz

Frequency = ____________________

☐2.1 ☐2.2

New Zealand Qualifications Authority

NATIONAL QUALIFICATIONS FRAMEWORK
MODERATION OF PHYSICS UNIT STANDARDS, 1997

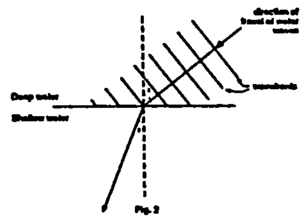
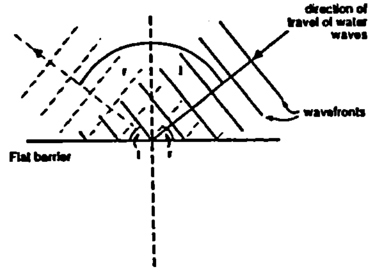
NATIONALLY PRESCRIBED ACTIVITY FOR PHYSICS UNIT STANDARDS LEVEL 2

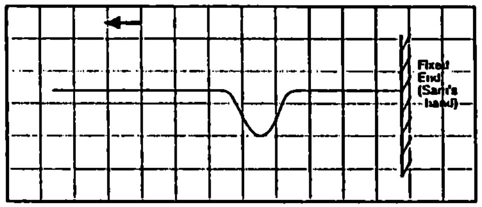
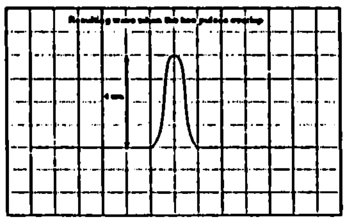
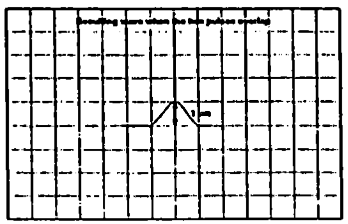
ASSESSMENT SCHEDULE

Unit Standard 6382		Demonstrate knowledge of waves.
Element	1.	Describe waves. Range: waves on springs and strings, water waves.
	2.	Determine an unkown quantity when a wave refracts. Range: two of – wave velocity, frequency, wavelength, period.

NATIONALLY PRESCRIBED ACTIVITY FOR PHYSICS UNIT STANDARDS LEVEL 2, 1997
UNIT STANDARD 6382
ASSESSMENT SCHEDULE

Assessment Schedule : Task 1

Task Number	Performance Criteria	Evidence	Judgement
(a) 1.	1.1	 Fig. 2	Angle of refraction (r) should clearly be less than the angle of incidence (i). Arrow should be drawn on the direction ray. Wavefronts are not necessary on the diagram. Angles do not need to be labelled.
(a) 2. (a) 3. (a) 4.	1.1 1.1 1.1	speed decreases wavelength decreases frequency stays the same	Equivalent statements accepted
(b) 1,2,3.	1.1	 Fig. 3	1. Arrow should be drawn on the direction ray. 2. Angle of incidence (i) and angle of reflection (r) must be labelled correctly. (Any of the values shown are acceptable). Angles must appear equal. 3. Wavefronts must be drawn at right angles to the ray (allow reasonable tolerance, say +/- 2°). At least 4 wavefronts need to be drawn.
(c) 1	2.1 2.2	$f = \frac{12}{4 \times 60}$ $f = 0.050$ Hz	All working should be shown. Answer should be correct. Answer should be quoted with the correct unit. Accept either Hz or cycles per second.
(c) 2	2.1	$v = f\lambda$ $v = 0.050 \times 5.0$	Formula should be stated. All working should be shown.

	2.2	$v = 0.25\ \mathrm{m\ s^{-1}}$	Answer should be quoted with correct unit. Answer consequential on (c)1. In both calculations accept the exact value or suitable rounding.
Assessment Schedule : Task 2			
(a) 1.	1.1		Reflected wave pulse should show 180° phase change. Reflected amplitude should be the same as the incident amplitude (+/-1 mm). Reflected wave should be the same shape as the incident wave. Reflected wave can be drawn at any position along the rope.
(a) 2.	1.1	Energy has been lost on reflection	Equivalent statement accepted.
(b) 1, 2.	1.2	1.  2. 	The shape of the resultant waveforms should resemble those drawn in the evidence statement with the correct amplitude and the correct phase. Amplitudes should be labelled as 4 cm or 1 cm or are the equivalent number of squares on the grid.

(c) 1	1.1	Amplitude is less. No phase change. Direction of travel is the same as incident pulse.	Equivalent statements accepted. Any two correct to achieve credit.
(c) 2	1.1	Amplitude is less. No phase change. Direction of travel is opposite to incident pulse.	Equivalent statements accepted. Any two correct to achieve credit.
Assessment Schedule : Task 3			
(a) 1, 2.	2.1	$n_1v_1 = n_2v_2$ $1.00 \times 3.00 \times 10^{8} = 1.33 \times v_2$	Formulae should be stated. All working should be shown.
	2.2	$v_2 = 2.26 \times 10^{8}$ m s^{-1} $n_1v_1 = n_2v_2$ $1.00 \times 3.00 \times 10^{8} = 1.34 \times v_2$ $v_2 = 2.24 \times 10^{8}$ m s^{-1}	Answers should be quoted with correct unit. Answers should be to 3 significant figures.
(b)	2.1	$v = f\lambda$ $3.00 \times 10^{8} = f \times 6.00 \times 10^{-7}$	Formula should be stated. All working should be shown.
	2.2	$f = 5.00 \times 10^{14}$ Hz	Answer should be quoted with correct unit.

ELEMENT 1

To gain credit for pc 1.1 — In Task 1(a) 1,2,3,4 - three out of four need to be correct and Task 2(c)1 needs to be correct.
Task 1(b) - two out of three need to be correct and either Task 2(a) or Task 2(c)2 need to be correct.

To gain credit for pc 1.2 — Either Task 2(b)1 or Task 2(b)2 needs to be correct.

ELEMENT 2

To gain credit — Task 1 (c)2, Task 3(a)1, Task 3(a)2 - one out of three need to be correct.
Either Task 1(c)1 or Task 3(b) needs to be correct.

Results of the 1997 end-point assessor judgement agreement trial.

Task	Performance criterion assessed	NM decision	Number of assessors who judged the student answer to meet the performance criterion	Number of assessors who judged the student answer not to meet the performance criterion	Percentage of end-point assessor judgements which agreed with the consensus decision
1a1	1.1	No	0	29	100
1a2	1.1	Yes	28	1	97
1a3	1.1	No	1	28	97
1a4	1.1	Yes	29	0	100
1b1	1.1	Yes	27	2	93
1b2	1.1	Yes	27	2	93
1b3	1.1	No*	13	16	55
1c1	2.1	Yes	29	0	100
	2.2	Yes	29	0	100
1c2	2.1	Yes	21	8	72
	2.2	Yes	26	3	90
2a1	1.1	Yes	29	0	100
2a2	1.1	Yes	29	0	100
2b1	1.2	No	0	29	100
2b2	1.2	Yes	29	0	100
2c1	1.1	Yes	29	0	100
2c2	1.1	Yes	25	4	86

3a1	2.1	Yes	26	3	90
	2.2	No	0	29	100
3a2	2.1	No	9	20	69
	2.2	No	0	29	100
3b	2.1	No	1	28	97
	2.2	No	0	29	100
Mean percentage of agreement of end-point assessor judgements					93
Standard deviation of percentages of agreement of end-point assessor judgements					12

For task 1b3 the national Moderator decision was that the student had not met performance criterion1.1 since the angle of incidence is not equal to the angle of reflection and the refracted ray does not meet the reflected wave fronts at right angles. This was not a clear decision since there is room for professional judgement.

Appendix 18

1996 Level 2 Physics Student Questionnaire

(For schools assessing against the Physics Unit Standards)

This questionnaire is part of a research project investigating the views of Sixth Form Physics students about their physics course and assessment programme.

Please answer the following questions by circling the appropriate answer or by writing your response in the spaces provided. The information you provide will remain confidential.

School: ______________________________

Gender: Male Female

General

1. How have you found your study of physics this year?

Very enjoyable	Enjoyable	Not sure	Unenjoyable	Very unenjoyable
5	4	3	2	1

2. State the major reason for your answer to Question 1.

____

____

____

3. How would you describe your workload in physics in comparison with the other Sixth Form subjects you are studying this year?

Far greater	Greater	Not sure	Less	Far less
5	4	3	2	1

Comment:

____

____

____

4. How would you describe the number of assessments (exams, tests, assignments, etc.) for physics in comparison with your other Sixth Form subjects?

Far greater	Greater	Not sure	Less	Far less
5	4	3	2	1

Comment:

____

____

____

Assessment against the Physics Unit Standards

5. How satisfied are you with the way you are assessed against the Physics Unit Standards?

Very satisfied	Satisfied	Not sure	Unsatisfied	Very unsatisfied
5	4	3	2	1

Comment:

____

____

____

6. How well do you understand the process by which credit is awarded for Physics Unit Standards?

Very well	Well	Not sure	Poorly	Very poorly
5	4	3	2	1

Comment:

____

____

____

____

7. How useful do you find the results of your Physics Unit Standard assessments for describing how well you performed on a particular task?

Very useful	Useful	Not sure	Of limited use	No use
5	4	3	2	1

Comment:

____

____

____

____

8. How useful do you find the written feedback you receive on your assessments against the Physics Unit Standard for describing how you can improve your performance?

Very useful	Useful	Not sure	Of limited use	No use
5	4	3	2	1

Comment:

____

____

____

Assessment for Sixth Form Certificate Physics

9. How satisfied are you with the way you are assessed for Sixth Form Certificate Physics?

Very satisfied	Satisfied	Not sure	Unsatisfied	Very unsatisfied
5	4	3	2	1

Comment:

____

____

____

10. How well do you understand the process by which Sixth Form Certificate grades are awarded in Physics?

Very well	Well	Not sure	Poorly	Very poorly
5	4	3	2	1

Comment:

---

11. How useful do you find the marks you receive on your assessments for Sixth Form Certificate for describing how well you performed on a particular task?

Very useful	Useful	Not sure	Of limited use	No use
5	4	3	2	1

Comment:

---

12. How useful do you find the marks you receive on your assessments for Sixth Form Certificate for describing how you can improve your performance?

Very useful	Useful	Not sure	Of limited use	No use
5	4	3	2	1

Comment:

Thank you for taking the time to complete this questionnaire.

Appendix 19

1996 Form 6 Physics Student Questionnaire

(For schools not assessing against the Physics Unit Standards)

This questionnaire is part of a research project investigating the views of Sixth Form Physics students about their physics course and assessment programme.

Please answer the following questions by circling the appropriate answer or by writing your response in the spaces provided. The information you provide will remain **confidential**.

School: ______________________________

Gender: Male Female

General

1. How have you found your study of physics this year?

Very enjoyable	Enjoyable	Not sure	Unenjoyable	Very unenjoyable
5	4	3	2	1

2. State the major reason for your answer to Question 1.

____

____

____

3. How would you describe your workload in physics in comparison with the other Sixth Form subjects you are studying this year?

Far greater	Greater	Not sure	Less	Far less
5	4	3	2	1

Comment:

____

____

____

4. How would you describe the number of assessments (exams, tests, assignments, etc.) for physics in comparison with your other Sixth Form subjects?

Far greater	Greater	Not sure	Less	Far less
5	4	3	2	1

Comment:

Assessment for Sixth Form Certificate Physics

5. How satisfied are you with the way you are assessed for Sixth Form Certificate Physics?

Very satisfied	Satisfied	Not sure	Unsatisfied	Very unsatisfied
5	4	3	2	1

Comment:

6. How well do you understand the process by which Sixth Form Certificate grades are awarded in Physics?

Very well	Well	Not sure	Poorly	Very poorly
5	4	3	2	1

Comment:

7. How useful do you find the marks you receive on your assessments for Sixth Form Certificate for describing how well you performed on a particular task?

Very useful	Useful	Not sure	Of limited use	No use
5	4	3	2	1

Comment:

8. How useful do you find the marks you receive on your assessments for Sixth Form Certificate for describing how you can improve your performance?

Very useful	Useful	Not sure	Of limited use	No use
5	4	3	2	1

Comment:

Thank you for taking the time to complete this questionnaire.

Appendix 20

1997 Form 6 Physics Student Questionnaire

This questionnaire is part of a research project investigating the views of Sixth Form Physics students about their physics course and assessment programme.

Instructions

Please answer the following questions by circling the appropriate answer or by writing your response in the spaces provided. The information you provide will remain confidential.

Section A is to be completed by **all** students

Section B is to be answered **only** by students who are being assessed using the Physics Unit Standards

Section A

To be answered by all students

School: ______________________________

Gender: Male Female

1. Which qualification(s) are you currently being assessed for?

Sixth Form Certificate

National Certificate (Unit Standards)

Both Sixth Form Certificate and National Certificate (Unit Standards)

2. How enjoyable have you found your study of physics this year?

Very enjoyable	Enjoyable	Not sure	Unenjoyable	Not at all enjoyable
5	4	3	2	1

3. State the main reason for your answer to Question 2.

____

____

____

4. How would you describe your workload in physics in comparison with the other Sixth Form subjects you are studying this year?

Far greater	Greater	Not sure	Less	Far less
5	4	3	2	1

Comment:______________________________________

____

____

____

5. How would you describe the number of assessments (exams, tests, assignments, etc.) for physics in comparison with your other Sixth Form subjects?

Far greater	Greater	Not sure	Less	Far less
5	4	3	2	1

Comment:______________________________________

____

____

____

6. How satisfied are you with the way you are assessed for Sixth Form Certificate Physics?

Very satisfied	Satisfied	Not sure	Unsatisfied	Very unsatisfied
5	4	3	2	1

Comment:______________________________________

____

____

____

7. How well do you understand the process by which Sixth Form Certificate grades are awarded in Physics?

Very well	Fairly well	Not sure	Poorly	Very poorly
5	4	3	2	1

Comment:____

8. How useful do you find the marks and feedback you receive on your assessments for Sixth Form Certificate for describing how well you performed on a particular task?

Very useful	Useful	Not sure	Of limited use	No use
5	4	3	2	1

Comment:____

9. How useful do you find the marks and feedback you receive on your assessments for Sixth Form Certificate for describing how you can improve your performance?

Very useful	Useful	Not sure	Of limited use	No use
5	4	3	2	1

Comment:____

Section B

To be answered only by students who are currently being assessed against the Physics Unit Standards.

10. Which qualification do you consider to be more useful to your needs?

Sixth Form Certificate	National Certificate (Unit standards)

Comment:____

11. How well have you been informed about the new National Certificate?

Very well	Well	Not sure	Poorly	Very poorly
5	4	3	2	1

Comment:____

12. How satisfied are you with the way you are assessed against the Physics Unit Standards?

Very satisfied	Satisfied	Not sure	Unsatisfied	Very unsatisfied
5	4	3	2	1

Comment:____

13. What do you like about assessment based on Unit Standards?

____

____

____

14. What don't you like about assessment based on Unit Standards?

____

____

____

15. How well do you understand the process by which credit is awarded for assessment based on Physics Unit Standards?

Very well	Well	Not sure	Poorly	Very poorly
5	4	3	2	1

Comment:_______________________________________

____

____

____

16. How useful do you find the results and feedback you receive on your assessments for the Physics Unit Standard for describing how well you performed on a particular task?

Very useful	Useful	Not sure	Of limited use	No use
5	4	3	2	1

Comment:_______________________________________

____

____

____

17. How useful do you find the results and feedback you receive on your assessments against the Physics Unit Standard for describing how you can improve your performance?

Very useful	Useful	Not sure	Of limited use	No use
5	4	3	2	1

Comment:____

____

____

____

Thank you for taking the time to complete this questionnaire.

Appendix 21

1998 Form 6 Physics Student Questionnaire

This questionnaire is part of a research project investigating the views of Sixth Form Physics students about their physics course and assessment programme.

Instructions

Please answer the following questions by circling the appropriate answer or by writing your response in the spaces provided. The information you provide will remain confidential.

Section A is to be completed by **all** students

Section B is to be answered **only** by students who are being assessed using the Physics Unit Standards

Section A

To be answered by all students

School: ______________________________

Gender: Male Female

1. Which qualification(s) are you currently being assessed for?

Sixth Form Certificate

National Certificate (Unit Standards)

Both Sixth Form Certificate and National Certificate (Unit Standards)

2. How enjoyable have you found your study of physics this year?

Very enjoyable	Enjoyable	Not sure	Unenjoyable	Not at all enjoyable
5	4	3	2	1

3. State the main reason for your answer to Question 2.

____

____

____

4. How would you describe your workload in physics in comparison with the other Sixth Form subjects you are studying this year?

Far greater	Greater	Not sure	Less	Far less
5	4	3	2	1

Comment:______________________________________

____

____

____

5. How would you describe the number of assessments (exams, tests, assignments, etc.) for physics in comparison with your other Sixth Form subjects?

Far greater	Greater	Not sure	Less	Far less
5	4	3	2	1

Comment:______________________________________

____

____

____

6. How satisfied are you with the way you are assessed for Sixth Form Certificate Physics?

Very satisfied	Satisfied	Not sure	Unsatisfied	Very unsatisfied
5	4	3	2	1

Comment:______________________________________

____

____

____

7. How well do you understand the process by which Sixth Form Certificate grades are awarded in Physics?

Very well	Fairly well	Not sure	Poorly	Very poorly
5	4	3	2	1

Comment:____

8. How useful do you find the marks and feedback you receive on your assessments for Sixth Form Certificate for describing how well you performed on a particular task?

Very useful	Useful	Not sure	Of limited use	No use
5	4	3	2	1

Comment:____

9. How useful do you find the marks and feedback you receive on your assessments for Sixth Form Certificate for describing how you can improve your performance?

Very useful	Useful	Not sure	Of limited use	No use
5	4	3	2	1

Comment:____

Section B

To be answered only by students who are currently being assessed against the Physics Unit Standards.

10. How satisfied are you with the way you are assessed against the Physics Unit Standards?

Very satisfied	Satisfied	Not sure	Unsatisfied	Very unsatisfied
5	4	3	2	1

Comment:____

____

____

____

11. How well do you understand the process by which credit is awarded for assessment based on Physics Unit Standards?

Very well	Well	Not sure	Poorly	Very poorly
5	4	3	2	1

Comment:____

____

____

____

12. How useful do you find the results and feedback you receive on your assessments for the Physics Unit Standard for describing how well you performed on a particular task?

Very useful	Useful	Not sure	Of limited use	No use
5	4	3	2	1

Comment:____

____

____

____

13. How useful do you find the results and feedback you receive on your assessments against the Physics Unit Standard for describing how you can improve your performance?

Very useful	Useful	Not sure	Of limited use	No use
5	4	3	2	1

Comment:____

14. How useful do you find the opportunity for reassessment?

Very useful	Useful	Not sure	Of limited use	No use
5	4	3	2	1

Comment:____

15. How well do you think Unit Standard credits are accepted by employers?

Very well	Well	Not sure	Poorly	Very poorly
5	4	3	2	1

Comment:____

16. How well do you think unit standard credits are accepted for entry into tertiary study?

Very well	Well	Not sure	Poorly	Very poorly
5	4	3	2	1

Comment:____

17. Which qualification do you consider to be more useful to your needs?

Sixth Form Certificate   National Certificate (Unit standards)

Comment:____

18. How well have you been informed about the new National Certificate?

Very well	Well	Not sure	Poorly	Very poorly
5	4	3	2	1

Comment:____

19. Indicate whether each of the following statements about the National Qualifications Framework is true or false by circling the T or F next to each statement.

a)	To be awarded the National Certificate in Educational Achievement you must complete a Polytech course.	T	F
b)	To be awarded a National Certificate you must complete a minimum of 240 credits	T	F

c)	There are a total of 10 physics unit standards at level 2 of the Qualifications Framework.	T	F
d)	All physics unit standards are worth the same number of credits.	T	F
e)	If you complete half of the total number of elements of a unit standard you get awarded half of the number of credits for that unit standard	T	F
f)	The Record of Learning is issued by the NZQA.	T	F
g)	You can only get credit for a unit standard if your work has been checked by a moderator.	T	F
h)	The physics unit standards you are assessed against in your Year 12 physics course are at level 2 of the Qualifications Framework.	T	F
i)	Each element you complete is worth one credit towards a National Certificate.	T	F
j)	To get entry into Form 7 you must do SFC in addition to unit standards.	T	F

20. What do you like about assessment based on Unit Standards?

____

____

____

21. What don't you like about assessment based on Unit Standards?

____

____

____

Thank you for taking the time to complete this questionnaire

Appendix 22

Schedules of questions used for the 1996-1998 case study interviews

The schedule of questions used for the 1996 interviews

The questions which were asked in 1996 related specifically to the trial and were formulated to establish baseline information for each school in the sample. All of the questions were repeated in subsequent years and changes in responses analysed to establish longitudinal patterns:

- What was the total number of students taking form 6 physics in 1996?
- What was the total number of sixth form physics classes in 1996?
- What was the total number of physics teachers at the school in 1996?
- Did you take part in the 3 day NZQA Unit Standard training programme?
- How useful did you find the 3 day NZQA Unit Standard training programme?
- What did you have to do to implement the Physics Unit Standards and were there any problems related to this?
- How have you coped with the recording and reporting aspects of assessing against Unit Standards?
- How has the introduction of the Physics Unit Standards affected your teaching?
- How has the introduction of the Physics Unit Standards affected student learning in your classes?
- How has the introduction of the Physics Unit Standards affected student motivation in your classes?
- How does student achievement in the physics Unit Standards (no of credits) compare with their Sixth Form Certificate grades?
- How useful do you find each of the following aspects of the physics moderation system?

    Moderation of assessment plan

    Moderation of assessment activities

    Moderation of assessor judgements (assessed student work)

- What do you see as the purpose of the nationally prescribed activity?
- Do you have any comments, about the way this was conducted in 1996, or about the 1996 NPA?
- How useful did you find the cluster meetings in 1996?

How could these have been improved?

- How useful did you find the assessment guide for physics in 1996?

How could this have been improved?

- Please comment on the accuracy, usefulness etc. of the NZQA communication you received last year.
- What are your plans for implementing the new curriculum?
- How did you go about the dual assessment for Unit Standards and SFC in 1996?
- What do you see as the preferred method of assessment for national qualifications at each of levels 1-3?
- How did the PPTA framework freeze affect your physics teaching and assessment this year?
- At what levels will you assess against the Physics Unit Standards in 1997?
- Please sum up your experiences with Unit Standards this year? Positives and negatives.

The schedule of questions used for the 1997 interviews

The schedule of questions which was used in 1997 consisted of all the questions asked in 1996 plus a number of additional questions.The additional questions were asked to follow up changes in the moderation system, Physics Unit Standards , government policy and the PPTA Framework Inquiry. In particular extra questions were added to investigate the schools responses to issues relating to authenticity of student work, reassessment, resubmission, sufficiency and the practicalities of portfolio keeping. The aim of the repeated questions was to investigate the longitudinal changes related to unit standard assessment and administration which occurred since the 1996 trial, as experienced by the teacher in charge of physics at each school in the sample of Canterbury schools. The interviews were conducted in term 4 of 1997.

- What was the total number of physics teachers at the school in 1997? Were there any staffing changes since last year?
- Are you currently assessing against the Physics Unit Standards? At what levels?
- Did the school assess against the level 1 Science Unit Standards in 1996? If so are these students continuing at level 2?
- Are students who did level 2 Physics Unit Standards in 1996 continuing at level 3?
- How does the workload this year compare with last year?
- How did you go about the dual assessment for Unit Standards and SFC in 1997? Did this pose any problems? Explain.
- Under what circumstances do you allow resubmission of student work?
- How do you organise reassessment? How many times is a student allowed to be reassessed?
- How do you decide what is sufficient evidence for a student to gain credit for an element?
- What measures do you take to ensure that work which is not completed under formal test conditions is authentic?
- Do you use student portfolios? How and where do you store these?
- How do you find the revised unit standards? Are they better than version 1?
- How useful do you find the revised Assessment Guide and packs of assessment activities?
- Do you think there is a need for the NPA?
- Do you have any comments about the way this was conducted in 1997 or about the
- Do you have any comments on :

    the proposals of the Green paper on senior secondary school assessment;

the PPTA Qualification Framework Inquiry findings?

- At what levels will you assess against the Physics Unit Standards in 1998?
- Please sum up your experiences with Unit Standards this year? Positives and negatives.
- What do you see as the main changes related to unit standard which have occurred since last year in terms of :

  your personal experience;

  student attitudes;

  the staff attitudes towards framework assessment at your school

  the moderation requirements?

The schedule of questions used for the 1998 interviews

The schedule of questions which was used in 1998 consisted of all the questions asked in 1996 and 1997 plus the additional questions listed below. The teacher interviewed was also supplied with a copy of the summary of the 1996 and 1997 interviews and asked if this factually represented the history of assessment against the Physics Unit Standards in their school. It was also an opportunity for them to elaborate on any issues and to comment on whether their views had changed in any way.

- At what levels did you assess against the Physics Unit Standards in 1998? What is your rationale for this?
- At what levels do you plan to assess against the Physics Unit Standards in 1999?
- Do you agree with the decision not to have a Nationally Prescribed Activity from 1998 onwards?
- What is your opinion about NZQA providing pre-moderated assessment activities?
- What do you think of the quality of the assessment activities provided in 1998 ie *The Physics of Sport* and *Atoms and Radiation*?

- What internal moderation procedures do you have in place within the school to ensure that different teachers assess to the same standard?
- Comment on your experience this year with

    the moderation of the assessment plan

    the moderation of assessment activities you submitted

    the verification of assessed work
- In your opinion, how many items of assessed student work should be submitted to the Local Moderator for verification of assessor judgements