NATIONAL STANDARDS: A SECONDARY ANALYSIS OF THE NZCER 2010 PRIMARY AND INTERMEDIATE SCHOOLS NATIONAL SURVEY

BY

JESS MAZENGARB

A thesis

submitted to the Victoria University of Wellington

in fulfilment of the requirements for the degree of

Master of Arts (Applied) in Social Science Research

Victoria University of Wellington

(2015)

Abstract

This thesis describes a secondary analysis of the NZCER 2010 Primary and Intermediate Schools National Survey. The aim of this project was to conduct an exploratory investigation into the attitudes of education stakeholders toward National Standards. Open response data from the NZCER 2010 Primary and Intermediate Schools National Survey principal, teacher, trustee, and parent questionnaires was coded in two different ways. First, a thematic framework was developed and the open responses were coded against the themes. The open responses were then recoded as a binary attitude variable, according to whether they were considered to convey a negative or positive attitude toward the standards. Logistic regression and Fisher's exact tests were used to determine statistically significant relationships between the binary attitude variable and other elements of the survey data, in the interests of finding predictors for attitude to the National Standards. These analyses were carried out with a selection of items from each of the principal, teacher, and trustee questionnaires; and with all of the items from the parent questionnaire. For teachers and principals (and to a lesser extent, trustees), findings from the thematic analysis of open responses largely reflected concerns about the standards being voiced by sector groups and academics around the time the survey was undertaken. Results from the statistical analysis of the binary attitude variable indicated that principal and teacher attitudes were associated with teaching experience and career plans. General findings from the analysis of parent responses included a suggestion that parents may lack knowledge and understanding of the National Standards, and may be expecting the standards to deliver something they are not designed for. Patterns in parent responses also indicated that attitudes to the National Standards might be related to wider perceptions of the purpose of education, and the various roles of education stakeholders. A similar pattern emerged in analysis of trustee responses. More specific findings related to parent attitude included associations with student year level, and with school reporting practice.

Acknowledgements

I am immensely grateful to my family and friends for their support. I would also especially like to thank my supervisor, John Haywood, for his incredible patience and his faith in me.

CONTENTS

1 INTRODUCTION 5

2 BACKGROUND TO THE NATIONAL STANDARDS 8

2.1 The International Context 8

2.2 The development of National Standards in New Zealand 13

3 THE NZCER 2010 PRIMARY AND INTERMEDIATE SCHOOLS NATIONAL SURVEY 17

3.1 The Tools 17

3.2 The Findings 18

4 SECONDARY ANALYSIS 28

4.1 Coding by theme – Methods and results 28

4.2 Creating an attitude measure from open response data 47

4.3 Considering whether to use weighting 56

4.4 Methods for statistical analysis 59

4.5 Analysis and results 67

5 DISCUSSION AND CONCLUSIONS 87

6 GLOSSARY 97

7 REFERENCES 98

APPENDIX 1: 2010 NZCER National Primary and Intermediate Survey - Questionnaire for Parents or Caregivers 106

TABLES

Table 1: Number of open responses by stakeholder type 43

Table 2: Full list of theme-based coding categories 44

1 INTRODUCTION

One of the National Party's core policies in the 2008 general election was the introduction of a set of National Standards in literacy and numeracy for students in Years 1 to 8. These standards were to consist of benchmarks for achievement for each year level that would indicate a student was "on track" for obtaining National Certificate of Educational Achievement level 2. Teachers would be required to make judgements about each student's progress toward (mid-year) and achievement against (anniversary/end of year) those benchmarks using a range of evidence in the form of work samples, assessment and observation (Ministry of Education, 2009 (1); Chamberlain, 2009). Judging students against standards would allow the identification of those needing targeted funding to raise achievement levels, and would also provide a framework for regular reporting of student achievement to parents in "plain language" (National Party, n.d.). The National Standards were developed by curriculum experts in mathematics and literacy in 2009 (Ministry of Education, 2010 (2); Ministry of Education, n.d. (1)), and were introduced for implementation in schools in 2010. Schools would be required to show progress against targets for achievement in their 2012 annual reports (Ministry of Education, 2009 (1)).

The National Standards were met with much concern from professionals and academics within the education sector, who pointed to growing acknowledgement of damaging effects arising from national testing systems overseas, and urged further work be done in development and trialling before such a significant policy move was undertaken (Thrupp et al., 2009). However, the Government was unbudging, and implementation of the standards went ahead as planned, despite continued vocal opposition, and heavily strained relationships.

We are now several years down the track from implementation, and regardless of any persistent disagreement over the soundness of the philosophical basis and practical implementation of the policy, it would seem that the most constructive way forward is to take whatever action possible to maximise the potential benefits and minimise the potential pitfalls of National Standards.

While many critics of National Standards point to perceived failure of similar systems overseas, New Zealand is a unique context and the framework for National Standards implementation includes aspects that set it apart from international examples, such as the use of Overall Teacher Judgements (OTJs) (Hattie, 2011). As Hattie suggests, this context provides an invaluable opportunity for research. If we are to make the most of National

Standards, the first step is to build a picture of what does and what could exist through the systematic collection of evidence.

Since 1987, the New Zealand Council for Educational Research (NZCER) has been conducting national surveys of primary and intermediate schools in New Zealand at generally 3-year intervals. The purpose of these surveys is to identify emerging issues and track trends and changes over time within the primary and intermediate education sector. The NZCER 2010 Primary and Intermediate Schools National Survey was the first of these surveys to be undertaken post introduction of National Standards. It included questions for teachers, principals, Board of Trustees members and parents about their experiences and expectations with regard to the standards and their implementation. A broad analysis of the survey data was completed and reported in December 2010 (Wylie & Hodgen, 2010). This analysis included frequency patterns and some cross-tabulations. In 2011, as part of a larger body of work, the data was used once more to carry out a more in-depth factor analysis of the teacher survey data, looking specifically at their previous experience of moderation. This analysis indicated a relationship between teachers' prior experience of moderation and their perception of a positive impact from National Standard: those teachers who had no or limited prior experience of working collegially to moderate student work were more likely to report positive changes associated with the introduction of National Standards. When interpreted in the context of the wider research into moderation, this suggests that for a subgroup of teachers, the introduction of National Standards may have led to positive new ways of working with assessment data (Hipkins & Hodgen, 2011).

This thesis describes a secondary analysis of the NZCER 2010 Primary and Intermediate Schools National Survey, designed to further explore the dataset for information that might help build a better understanding of the National Standards and their implications for stakeholders in the education context. The analysis takes an exploratory approach, using established statistical techniques to look for trends and relationships that might underlie the views and attitudes toward National Standards held by those within the stakeholder groups surveyed.

Open responses on National Standards were chosen as the basis for the secondary analysis. All of the surveys included an opportunity to comment on National Standards, making it possible to access views from teachers, principals, trustees, and parents; and to make some comparison between these groups. Two alternative coding systems were applied to open responses. The first involved a thematic analysis, with the original intention being to use

statistical methods to look for relationships between themes identified in stakeholder comments and other variables extracted from the remainder of the survey data. However, the results of thematic coding precluded such an approach, due to the subjectivity involved in interpretation, and the large number of resultant categories with relatively few responses in each. Instead, quantitative analysis was limited to the production of frequencies, and themes were treated as illustrative. Open responses were then recoded according to whether they indicated a negative or a positive view of the National Standards, creating a new attitude variable for each stakeholder group. This coding framework was tested for reproducibility using a second coder and measuring agreement according to several different inter-rater reliability statistics. Agreement was found to be relatively strong overall.

Weighting of data to improve the representativity of the respondent sample according to school-based characteristics was considered, but rejected on the basis that respondents providing an open response constituted a self-selected sample, and as such, representativity was not an appropriate goal.

Potential relationships between the newly created attitude variable and other variables contained within the survey dataset were explored using logistic regression and Fisher's exact tests, with variable reduction in the form of exploratory factor analysis applied to some banks of associated Likert-type items.

Some background to the development of National Standards in New Zealand is provided in chapter 2, and the findings of the initial analysis of the NZCER 2010 Primary and Intermediate Schools National Survey are summarised in chapter 3. Methods and results are described in detail in chapter 4. Results are then discussed in chapter 5, in the context of other relevant research and literature.

2 BACKGROUND TO THE NATIONAL STANDARDS

2.1 The International Context

Increasing accountability within the education sector, and setting and raising standards of performance, have been common global themes in education reform in recent decades. These themes have been conceptualised and explained in reference to the broader social, political, and economic context by a range of commentators. While the subject of global education reform is one that has attracted much attention, and for the most part, the complexities of purpose and process are beyond the scope of this thesis, some approaches to understanding education reform as an international phenomenon are presented briefly here in order to provide some context for recent policy development in New Zealand.

Globalisation and Human Capital

The economics of education developed as a field in its own right in the 1960s. According to Machin (2008), its origins can be traced to Becker's *Human Capital*, published in the United States in 1964. The human capital approach is based on the idea that additional investment in human capital, through education and healthcare, leads to increased labour productivity. The 1960s was also when the International Association for the Evaluation of Educational Achievement (IEA) began undertaking research with the aim of comparing quality of education across different systems. Until the 1980s, the results of the IEA studies were mainly used by researchers and educators. However, from the 1990s, governments started to become more interested in cross-national comparisons of education systems (Ross et al., 2006). This burgeoning interest can be attributed, at least in part, to the growing body of research linking primary and secondary education to economic development that stemmed from the human capital approach (Sahlberg, 2006) and the increasing prominence of economists in educational theory and research (Guthrie & Pierce, 1990). As education came to be seen as closely linked not only with an individual's ability to earn and participate in society, but with a whole society's social and economic mobility (Ross et al., 2006), the international competitiveness of a nation's education system took on a new level of meaning and significance. Cross-national comparisons of education quality were able to be seen as a sort of proxy measure for economic competitiveness.

Changing views about the role and purpose of education are reflected in the United Nations Educational, Scientific and Cultural Organization (UNESCO) Education for All Movement.

The 2000 Dakar Framework for Action, adopted by the participants in the World Education Forum, 2000 stated that those governments and organisations were committed to "improving all aspects of the quality of education and ensuring excellence of all so that recognized and measurable learning outcomes are achieved by all, especially in literacy, numeracy and essential life skills." (UNESCO, 2000). The focus on quality arises from concern that participation in education, in and of itself, is not sufficient to ensure that students acquire the necessary knowledge and skill for participation in the global economy, and therefore governments must monitor outcomes to ensure education systems are "adequate to meet their economic goals" (Kellaghan, 2006, p.52). This shift is described by Machin (2008) in the context of human capital theory as a change in the classification of education from a consumption good to an investment good.

The most widely known cross-national comparison of education quality, and perhaps the most significant monitoring programme in education, is the OECD's Programme for International Student Assessment (PISA). PISA was first undertaken in 2000 and has been carried out every three years since. During each cycle, a sample of 15-year-old students from each of the participating countries sit tests in reading, mathematics, and science, with the intention of assessing "the extent to which students near the end of compulsory education have acquired some of the knowledge and skills that are essential for full participation in modern societies" (OECD, 2010, p.17). This intention is consistent with the idea of education as an investment in human capital – measuring the knowledge and skill level people are bringing to the labour force. All OECD countries participate in PISA, together with a growing number of partner countries – in 2000, 43 countries took part, while in 2012 there were 65 participating countries. Seventy-one countries are "signed up" to participate in PISA 2015 (OECD, n.d.). The results of the PISA assessments are used to rank countries according to performance.

PISA's significance in the monitoring and evaluation of education arises from its explicit role as a policy driver, as stated in the Foreword to *PISA 2009 Results: What Makes a School Successful?*:

"Most countries monitor students' learning and the performance of schools. But in a global economy, the yardstick for success is no longer improvement by national standards alone, but how education systems perform internationally. The OECD has taken up that challenge by developing PISA, the Programme for International Student Assessment, which evaluates the

quality, equity and efficiency of school systems in some 70 countries that, together, make up nine tenths of the world economy. PISA represents a commitment by governments to monitor the outcomes of education systems regularly within an internationally agreed framework and it provides a basis for international collaboration in defining and implementing educational policies."

(OECD, 2010, p.3).

This policy orientation sets PISA apart from other international comparison studies (Figazzolo, 2009), and cements the role of the OECD as one of the most important influencers of educational policy in the world today (Bieber & Martens, 2011). While the reaction to PISA results, and the convergence of policy with the resultant OECD recommendations vary substantially from one country to another, PISA contributes to the globalisation of policy through facilitating education "borrowing" and "lending" (Sahlberg, 2006), and creating additional pressure for performance. Given that evidence about how to improve the quality of education is lacking (note that the policy recommendations contained within PISA reports are based on factors found to be positively correlated with performance, but no causal relationships are established or claimed), many governments look to the educational policies and practices of countries they perceive to be more "successful", or to those reforms legitimated by international agencies such as the OECD when trying to improve education (Tatto, 2007).

While there is some debate about the soundness of the PISA methodology (Goldstein, 2004; Kreiner, 2011; Adams, 2011), the pressure to conform or be left behind has led policy makers to pay close attention to results and subsequent recommendations. Such pressure is encapsulated in the Foreword to *PISA 2009 Results: What Makes a School Successful?*:

"While better educational outcomes are a strong predictor of economic growth, wealth and spending on education alone are no guarantee for better educational outcomes. Overall, PISA shows that an image of a world divided neatly into rich and well-educated countries and poor and badly-educated countries is out of date.

This finding represents both a warning and an opportunity. It is a warning to advanced economies that they cannot take for granted that they will forever have "human capital" superior to that in other parts of the world. At a time of intensified global competition, they

will need to work hard to maintain a knowledge and skill base that keeps up with changing demands."

(OECD, 2010, p.3).

Accountability

As education has taken on responsibility for economic growth within a competitive and globalised world, it has become a high profile issue, with many stakeholders (Pigozzi, 2006) demanding information and accountability. In discussing what he refers to as "the emerging 'age of accountability'" (p.418), Hopmann (2008) suggests that while theories and models of accountability vary across contexts, there is general agreement that accountability procedures are becoming more pervasive, at least in Western societies. This pervasiveness applies to all areas of the public sector, and affects the relationship between the public and private sectors. These changing relationships involve significant redistribution of resources and responsibilities. Mons (2009) also points to changing relationships as a core feature of accountability developments, emphasising the growing importance in education of accountability to parents, and the general public. Mons notes that while performance-based accountability is not new, it has never been so pervasive. PISA provides an excellent example of system level accountability (Ross et al., 2006), allowing public scrutiny of the performance of education systems, through the publication of rankings on a global scale.

When it comes to the why and the how of the emergence of Hopmann's "age of accountability" (2008, p.418), different features are emphasised by different authors, in accordance with their disciplinary background or research emphasis. Machin (2008) describes the development of an "evaluation culture" in education arising from its increasing public profile. Machin links this development with significant growth in the economics of education as a research field, with econometrics of programme evaluation being employed to aid the creation of evidence-based policy. Sahlberg (2006) also looks to economics for explanation, suggesting that the drive for accountability is based on the assumption that the application of marketisation principles is the best way to improve education. Here, accountability is part of a suite of market values including productivity, effectiveness and, of course, competitiveness, aimed at improving performance. Sahlberg suggests that this process of marketisation of education has been at the core of global education reform since the 1990s.

Standards

The term standards is used in different and overlapping ways across multiple contexts in education. In the context of curriculum and assessment reform, standards are used to refer to content and achievement. Content standards generally apply to schools and education systems and specify the knowledge and skills that are expected to be taught. Achievement standards apply to students, and describe what is expected to be learned (Klenowski & Wyatt-Smith, 2010).

As already discussed, human capital theory has had a significant impact on education reform in recent decades. Many reforms are associated with a desire to ensure that the quality and content of education is satisfactory to meet the needs of the global economic environment and thereby maximise international economic competitiveness (Tatto, 2007). Given the importance of this investment in human capital, governments assume responsibility for articulating explicit expectations about what students should know and be able to do as a result of their participation in the education system. Active participation in a knowledge-based economy requires a minimum level of skill in reading, writing, mathematics, and science. Encouraging mastery of these basic literacies through setting learning targets for students and teachers has become common-place. This focus is reflected in, and reinforced by, international comparisons and benchmarking (Sahlberg, 2006) through projects like The International Mathematics and Science Study (TIMSS) and PISA.

Once expectations have been communicated using prescribed curricula and achievement levels, assessment is needed to measure whether or not students are achieving the desired standard of knowledge and skill acquisition (Tatto, 2007). As well as measuring outcomes, assessment works to reinforce the desired outcomes by influencing what is taught and what is learned (Torrance, 2011) by teachers and students concerned with meeting expectations to avoid negative sanctions or achieve positive ones (Guthrie & Pierce, 1990). An outcome-focused education system concerned with developing quality human resources needs an assessment system that is fit for purpose. Increasingly, traditional norm-referenced assessment is being seen as inadequate in providing the depth and breadth of information required by stakeholders (students, teachers, parents, potential employers, and the government). Instead, criterion-referenced assessment is being utilised for its ability to provide a fuller picture of what students can do, and what schools are producing (Torrance, 2011).

According to Lachat (1999), standards are about moving from ideals of access to education for all students, to quality education for all students. Common goals, formalised as standards, promote common learning outcomes (Sahlberg, 2006). In this way, standards-based reform has been promoted in the interests of equity (Lachat, 1999). Standards are also presented by reform proponents as a means of lifting achievement. The idea is that standards describing high expectations for all students, which are supported by integrated assessment, help and encourage students, teachers, and principals to focus on attaining those goals (Sahlberg, 2006; Klenowski & Wyatt-Smith, 2010). PISA has linked the presence of education standards to highly performing systems, particularly when combined with external accountability (Figazzollo, 2009; OECD, 2010).

2.2 The development of National Standards in New Zealand

National Standards were a key plank in the National Party's policy platform in the lead up to the 2008 general election. Citing New Zealand's “tail of under-achievement” (being the approximately 20 percent of New Zealand students who leave school without formal qualifications), the National Party stated a need to raise achievement through assessment of students in Years 1 to 8 against a nationally consistent set of performance benchmarks in literacy and numeracy that measured whether students were “on track” to achieving NCEA Level 2. Reference was also made to Education Review Office findings that many schools were not collecting and using assessment data effectively, or reporting adequately to parents; and, research by John Hattie concluding that schools often inflate student performance when reporting to parents (National Party, 2008; National Party, n.d.).

The National Standards were expected to improve achievement through creating a set of shared, high expectations for the achievement of all children; allowing teachers and schools to identify students at risk of underachieving, in order that early intervention could occur; ensuring that parents received information about their child’s achievement and next learning steps so that they could support learning at home; and, providing information for the Ministry of Education about areas of need so that resources could be targeted effectively (National Party, 2008; National Party, n.d.). While primary and intermediate schools were already required, under the National Education Guidelines, to assess and report on student progress and achievement in relation to the New Zealand Curriculum (Ministry of Education, 2010 (1)), there was little in the way of any specific requirement about what form that assessment and reporting should take. National Standards were intended, by providing a universal system

of benchmarking, to ensure that assessment measures in reading, writing, and mathematics would be comparable across time, between teachers, and across schools and that those measures would be made available to parents in order that they could see how their child's achievement compared with expectations for achievement relative to school year level, and with national patterns of achievement (National Party, 2008).

The National Party was elected to Government in November 2008, and the Education (National Standards) Amendment Act 2008 was passed in mid-December, giving the Minister of Education power to set National Standards in numeracy and literacy. According to a Beehive press release that preceded the passing of the Act (Tolley, 2008), the standards themselves were to be negotiated with the sector, through consultation about design, benchmarks for assessment tools, and forms of reporting to stakeholders. Consultation was to occur in 2009, and the standards published later that year. Schools were to begin reporting achievement against the standards to parents from 2010, to include school-wide targets for achievement against the standards in their 2011 charters, and report on progress and achievement against those targets in their 2012 annual reports.

The National Standards were developed during 2009 by literacy and numeracy working groups within the Ministry of Education. The standards were established primarily on the basis of the New Zealand Curriculum and the Literacy Learning Progressions. In determining appropriate expectations at each year level, a number of factors were considered – in order for standards to lift achievement, expectations should be high but reflect what students could be reasonably expected to achieve with quality instruction (i.e., be set at a level that not all students would reach when the standards were first introduced, with the expectation that the proportion of students reaching the standards would increase over time as a result of the improvements in teaching and learning that the standards would facilitate). In setting the mathematics standards, data from the Numeracy Development Projects, National Education Monitoring Project, Assessment Tools for Teaching and Learning, and Progressive Achievement Test were used to gauge high but reasonable achievement benchmarks. The focus of the reading and writing standards was to be the level of literacy expertise needed in order for students to engage successfully with all learning areas of the New Zealand Curriculum (Ministry of Education, n.d. (1); Ministry of Education, 2010 (2)). All of the standards should reflect a level of skill and knowledge that would indicate a student was on track for achieving NCEA Level 2 in Year 12 (Chamberlain, 2009).

As part of the implementation of National Standards, the National Administration Guidelines were amended to include requirements for reporting in relation to National Standards to students, parents, and the Ministry of Education. Under National Administration Guideline (NAG) 2A, schools are required to report to students and their parents twice yearly, in writing, on the student's progress and achievement against the standards. This reporting involves one or more interim judgements, as well as a final judgement. The interim judgement/s should address progress toward the relevant standards, while the final judgement is about determining whether expected standards have been met. For students in Years 1-3, expected standards relate to achievement after the student has been at school for one, two, or three years – being four, eight, or twelve terms. For students in Years 4-8, expectations relate to achievement at the end of the school year (Ministry of Education, n.d. (2)). NAG 2A also requires schools to report to the Ministry of Education on final judgements at the school level. Specifically, they must report numbers and proportions of students achieving "above", "at", "below", and "well below" National Standards by ethnicity, gender, and year level (Ministry of Education, 2013). Reporting to parents does not need to utilise the same terminology ("above", "at", "below", and "well below"), but must clearly convey the student's achievement in relation to the standards (Ministry of Education, n.d. (3)).

In determining whether students are meeting the expectations defined by the National Standards in reading, writing, and mathematics, teachers need to draw on a range of sources in order to make their professional judgement (an Overall Teacher Judgement, or OTJ). To assist teachers in making professional judgements, the literacy and numeracy working groups developed a set of illustrations to demonstrate what it might look like when a student is meeting the standard (Ministry of Education, n.d. (1); Ministry of Education, 2010 (2)). In making judgements about achievement against the National Standards benchmarks, teachers are expected to use observation, conversation (with the student), and results from formal assessments including standardised tests (Ministry of Education, n.d. (4)). No new tests were introduced with the National Standards and no specific tests were mandated for use in making judgements in order to avoid or mitigate the risk of excessive focus on test performance at the expense of quality teaching and learning (Ministry of Education, 2009 (2)).

To ensure the reliability of judgements, teachers are expected to engage in moderation – professional discussion to clarify expectations and assessment methods. Schools are required to design their own moderation processes (to suit their particular needs and circumstances),

which might involve teachers within a group, across a school, or from different schools (within a cluster of schools for instance) (Ministry of Education, 2009 (3)).

The introduction of National Standards generated much controversy and debate. The announcement in late 2009 that the Government was going to restrict funding for school support services to numeracy and literacy in line with the National Standards acted as a catalyst to bring debate over National Standards into the public sphere. Increasing concern and opposition was voiced by representatives of a range of key organisations in the education sector including the New Zealand Principals' Association, the Post Primary Teachers' Association and the New Zealand Educational Institute. There was also significant concern expressed by a number of prominent academics in education including Professor John Hattie who, although one of the founding architects of the National Standards concept, was unhappy about the direction the policy had taken. The debate became increasingly hostile, with some schools promising to boycott National Standards and the Minister of Education mentioning the possibility of sacking Boards of Trustees that refused to implement the standards. With unions being amongst those most outspoken against National Standards, the Prime Minister and the Minister of Education used a recent Education Review Office report to suggest that these unions were attempting to protect 30% of teachers who were under-performing (Kay, 2010; Clark, 2010). This was the political environment that provided the backdrop to the NZCER 2010 Primary and Intermediate Schools National Survey.

3 THE NZCER 2010 PRIMARY AND INTERMEDIATE SCHOOLS NATIONAL SURVEY

Since 1989, the New Zealand Council for Educational Research (NZCER) has been regularly surveying primary and intermediate schools in order to explore the impact of educational reform over time. The surveys are generally run every three years (NZCER, n.d.). The NZCER 2010 Primary and Intermediate Schools National Survey was the first iteration following the introduction of National Standards, and as such, the first opportunity for the survey series to explore the impact of National Standards on primary and intermediate schools. The survey was conducted in July 2010.

3.1 The Tools

The survey consisted of four questionnaires – one for principals, one for teachers, one for Board of Trustees members, and one for parents. The principal, teacher, and trustee questionnaires were divided into a number of themes relevant to those stakeholders. The themes varied across the different questionnaires, but National Standards was a theme common to all three.

Under the theme of National Standards, the principal questionnaire asked about what work had been done so far toward implementing National Standards, and the principal's experience of the introduction of the standards (including the available guidance and support and whether the standards themselves seemed robust). It also asked for the principal's opinion on the likely short term impact of National Standards within their school. The principal questionnaire included an additional question under another theme about the impact of the National Standards on the school's developmental work on the New Zealand Curriculum.

The teacher questionnaire asked whether the school had started implementing National Standards, and, if so, what changes teachers were seeing in their work as a result. The questionnaire contained a number of questions related to teachers' experience of making and moderating Overall Teacher Judgements (OTJs) and two sets of items, almost identical to those in the principal questionnaire, asking about teachers' experience of the introduction of the standards and their opinion on the likely short term impact of National Standards within their school. National Standards also featured under the professional learning and support theme in the teacher questionnaire.

The National Standards section of the trustee questionnaire was shorter than that of the principal or teacher, with just one set of items encompassing the Board's understanding of National Standards and experience of their implementation, and the board member's opinion of the likely impact of National Standards. However, National Standards were also mentioned in several other parts of the questionnaire. In the section concerned with the Board's capacity, achievements and issues, Board members were asked to indicate how much time they spent on National Standards relative to other tasks, and to indicate whether they considered National Standards to be a major issue facing the school. National Standards were also mentioned in sections on contact with parents and community engagement.

The parent questionnaire was not divided into themes, but did contain several questions about the nature and quality of student reports received from the school, one of which made specific reference to National Standards. Parents were also asked to indicate, from a list of school activities, which they had been involved in that year. One of the activities specified was an information session on National Standards. The parent questionnaire has been included as Appendix 1 to this document.

Questions were in multiple formats. Many were structured as Likert-type items where the respondent was asked to indicate their answer on a scale (e.g., strongly disagree to strongly agree, very important to not at all important, etc.). The scales varied, with 3-, 4-, and 5-point scales used in the questionnaires. Some questions provided a number of possible answers and asked respondents to tick all that apply, while others asked them to select only one response. There were also a number of places with space for respondents to provide open responses to questions. All four of the questionnaires contained an open-response box for comments on National Standards.

3.2 The Findings

At the end of 2010, NZCER published a report entitled *NZCER 2010 Primary and Intermediate Schools National Survey: A snapshot of overall patterns and findings related to National Standards – revised version* (Wylie & Hodgen, 2010). The data presented here are drawn from that report.

The NZCER 2010 Primary and Intermediate Schools National Survey was sent to a random sample of 350 schools in July 2010. Each of the schools received a survey for the principal, surveys for the chairperson and one other member of the Board of Trustees, and surveys for

half of the teaching staff. A cross-section of 35 of the schools also received surveys for a sample of one in seven parents. Completed surveys were received from 210 principals, 970 teachers, 257 trustees, and 550 parents. The highest response rate was from principals and the lowest from parents. Responses were weighted for most of the analysis (excluding cross tabulations) in the interests of representativity by decile and roll size.

Ninety-one percent of principals, and 83% of teachers, indicated that their school had begun to implement National Standards. Of those teachers and principals, the vast majority disagreed that timeframes and support for implementation had been sufficient. Most teachers indicated that they felt it was difficult for parents to understand the National Standards, and principals similarly indicated that they felt it was difficult for Boards of Trustees. Both teachers and principals largely agreed about the importance of consistency in, and moderation of, OTJs and felt that the Ministry of Education should be supporting schools in understanding OTJs, as well as supporting them to work together with other schools to moderate OTJs. Responses to questions about clarity and consistency of information and the robustness of the standards themselves showed more variation, but tended to be more negative than positive about the standards.

Of the teachers whose schools had started implementing National Standards, 49% thought they had discussed the National Standards in enough depth to understand them. Having discussed the standards in enough depth to understand them was generally associated with more positive views, both in relation to the clarity and consistency of information from the Ministry of Education, and the robustness of the standards.

Thirty-seven percent of teachers reported having had external (to the school) professional development funded by the Ministry of Education, with 69% of those teachers indicating it had little or no impact on their practice. A similar proportion reported taking part in whole-school professional development run by an external advisor, with 60% of those teachers indicating it had little or no impact on their practice.

Most teachers in schools which had begun implementing National Standards reported some experience in making OTJs. Thirty-eight percent had practised making OTJs. Twenty-four percent reported that their school had developed a chart showing how different assessments lined up with standards (there was no apparent relationship between the presence of such a chart and reported ease of making OTJs or time taken to make OTJs). Fifty-three percent had made and used OTJs in mid-year reports to parents. Those who had used OTJs in mid-year

reporting spent an hour more a week on their work, on average, than those who hadn't. Of the teachers with some experience in making OTJs, 64% indicated that it was clear to them where all their students were in relation to reading, and 63% said it was clear where all their students were in relation to mathematics. Only 49% felt it was clear where all their students were in writing. Similarly, a greater proportion of teachers reported difficulty putting together sources of evidence to make OTJs in writing (24%), than in mathematics or reading (each about 17%). Teachers who reported difficulty in putting together sources of evidence, or who felt it wasn't clear where all their students were in relation to the standards, were more likely to be from schools in which the standards were still being discussed to ensure they were understood. Almost all teachers (95%) reported using four or more sources of evidence to make OTJs.

Most teachers in schools that had started implementing National Standards reported previous experience in using benchmarks or progressions to assess students' progress. Previous experience moderating judgements against benchmarks in writing was reported by 77% of teachers, while 55% reported previous moderation in reading, and 53% reported previous moderation in mathematics. Of the teachers in schools that had started implementing National Standards, 24% said their school was not moderating OTJs; 49% said they had worked together to make OTJs for a range of students and discussed results for a consistent approach; 30% said they had moderated all OTJs with another teacher; and, 53% reported moderating by discussing borderline cases with another teacher or teachers. Moderation was more likely for writing than reading or mathematics and also more likely within than across year levels. Of the teachers who were making OTJs, 37% thought there was a high level of consistency in OTJs across the school, and 37% were unsure. Teachers who thought there was consistency were more likely to report that their school was moderating OTJs. Previous experience of moderation against benchmarks within the school was also associated with perceived consistency in OTJs.

In schools that had begun implementing National Standards, 42% of teachers indicated that they were spending more time on assessment as a result of National Standards. Thirteen percent felt they were gaining more insight into their students' learning needs than from assessment used in 2009. Forty-two percent reported working more with students to set goals based on assessment results.

Few teachers reported spending more time on reading, writing, or mathematics as a result of the introduction of National Standards, but 43% of those in schools that had begun implementation felt that they were spending less time on other curriculum areas. Spending less time on curriculum areas other than reading, writing and mathematics and spending more time on assessment were both associated with low morale, greater workload, greater work-related stress, and decreased sustainability of workload.

Some principals indicated that the introduction of National Standards had affected developmental work on the New Zealand Curriculum (NZC) within the school. Thirty-four percent reported that advisory support for NZC was difficult to access, and 34% reported that work on NZC had been cut back. Twenty-three percent of principals reported that working on the National Standards had helped the school's NZC developmental work in reading, writing, and mathematics.

In schools that had started implementing National Standards, 62% of principals said they used reading OTJs in reporting to parents, while 60% reported using maths OTJs and 60% reported using writing OTJs. In schools where implementation was underway, 47% of teachers reported spending more time on mid-year reporting to parents as a result of the standards.

Two thirds of parent respondents indicated that they had received a written report about their child's progress at mid-year. Fifty-three percent said they had discussed their child's progress with their child and their child's teacher at a set time, and 32% reported having discussed their child's progress with the teacher only at a set time. Seventy-two percent of parents felt they had received clear information on their child's overall progress. Two out of the 35 schools in the parent sample had not started implementing National Standards, according to the principal. Parents from those schools were just as likely as parents from the schools where implementation had begun to report having received clear information on their child's progress, and were also just as likely to report having received clear information about where their children were in relation to National Standards. Parents in the two schools that had not begun to implement the standards were more likely to have received a written report, but less likely to indicate that they had received clear information about their child's learning goals for the rest of the year, or helpful ideas to support their child's learning.

All parents were much less likely to indicate that they had received clear information about their child's progress in science, than they were to indicate that they had received clear information about their child's progress in reading, writing or mathematics.

Sixteen percent of parents reported that they had attended an information session on the National Standards at their child's school.

Eighty-five percent of principals and 86% of teachers thought that National Standards would not change patterns of achievement much because schools already identified student needs and worked hard to increase rates of progress. Sixty-five percent of principals, and 67% of teachers, thought that the standards wouldn't have much effect on achievement because schools needed additional support to change rates of learning progress. Seventy percent of principals and 61% of teachers indicated that they would see an increase in workload for little gain. Sixty-eight percent of principals and 71% of teachers felt that some parents were more anxious as a result of the introduction of National Standards. Seventy-one percent of principals and 67% of teachers reported that the likely impact of National Standards would depend on whether student performance was the main measure of school performance. Similar proportions (71% of principals and 65% of teachers) said the impact would depend on whether schools were able to keep National Standards in perspective, or integrate them into the school programme. Thirteen percent of teachers and 5% of principals agreed that parents of students identified as “below” or “well below” the standard would be more engaged in their child's learning in positive ways. Eleven percent of principals and 13% of teachers agreed that their practice would change because assessment against National Standards had indicated that their students were not doing as well as expected in relation to the standards. Eighteen percent of principals and 27% of teachers thought that National Standards had provided better data for decision making at the classroom level, while 19% of principals and 29% of teachers thought that National Standards had provided better data for decision making at the school level.

Seventy-four percent of trustee respondents said that their Board had discussed National Standards in depth, and 76% indicated that their Board had a clear picture of how National Standards were being implemented in the school. However, only 44% indicated that their Board had a really good understanding of National Standards. Twenty-eight percent of trustees thought National Standards were difficult to put into practice and 51% thought that school comparisons based on National Standards would unfairly damage the reputations of

some schools. Thirteen percent of trustees agreed that the timeframe for introduction of the standards was realistic, and 13% agreed that National Standards would definitely improve achievement in their school. Twenty-seven percent of trustees agreed that National Standards would increase parent engagement in a positive way.

Seventeen percent of trustees said they had taken part in a school meeting with parents about National Standards.

3.2.1 Cross-tabulations by school decile, school size, and year level

Cross-tabulations were used to identify differences in response by school or respondent characteristics. Results of cross-tabulations presented in Wylie and Hodgen's (2010) report seem to indicate some patterns associated with decile, size, and teaching year-level, which are summarised briefly below. School deciles are grouped as low (1-3), mid (4-7), and high (8-10). School sizes are defined as small (100 or fewer students), small-medium (101-200 students), medium-large (201-350 students), and large (351 or more students).

School decile

Principals of low-decile schools were generally more positive about National Standards than principals of higher decile schools. They (principals of low-decile schools) were also more likely to think Ministry of Education material had been clear and consistent, that they had had enough guidance and advice to be confident about the school's work on the National Standards, and that the time frame for implementation had been sufficient. Teachers in low-decile schools were more likely to say that they had had enough guidance and advice to feel confident about the school's work on National Standards (28% compared with 15% of teachers in mid- and high-decile schools), that the Ministry of Education guidance on reporting to parents had been clear and consistent (38% compared with 22% of teachers in mid- and high-decile schools), and that they had had enough time to make sense of National Standards before using them (14% compared with 6% of teachers in mid- and high-decile schools).

Low-decile schools were less likely to have begun implementing National Standards at the time of the survey. High-decile schools were more likely to have decided what sources of evidence to use to make OTJs, to have used OTJs in mid-year reports to parents, and to have moderated OTJs. Teachers from low-decile schools were more likely to have practised making OTJs.

High-decile schools were less likely than low-decile schools to be using the terms “below” and “well below” in reporting to parents (16% versus 36%). School decile was also associated with an increase in time spent on reporting to parents – the higher the decile the more likely teachers were to agree that they were spending more time reporting to parents as a result of National Standards.

Principals in low-decile schools were more likely to have used National Standards in planning, including as a means to identify students at risk (44% compared with 38% of all principals). However, they were also more likely to indicate that National Standards would not greatly change patterns of achievement in their school because they already identified student needs and worked hard to increase rates of progress. Teachers in decile 1-4 schools were more likely to report that National Standards would change their practice because students were not doing as well as expected (18% compared with 11% of teachers in decile 5-10 schools). Teachers in decile 1-2 schools were less likely to say that parents were more anxious as a result of National Standards (62% compared with 74% of teachers in decile 3-10 schools), and also less likely to indicate that National Standards would result in more workload for little gain (45% compared with 63% of teachers in decile 3-10 schools).

Trustees in low-decile schools were less likely to think it was difficult to put National Standards in place (17% compared with 39% of trustees in high-decile schools), and more likely to think that National Standards would definitely improve student achievement levels at their school (32% compared with 15% of trustees in mid-decile schools, and 10% of trustees in high-decile schools). However, they were more likely to agree that the school would not have the resources to support all students identified as being below standard unless they were to cut spending in other areas, or get more government support (58% compared with 35% of trustees in high-decile schools).

School size

Principals of small schools were more likely to be positive about National Standards than principals of larger schools. They were more likely to report that their work on National Standards was having a positive impact on their NZC development work (33% compared with 19% of principals of large schools), and more likely to report positive impacts in terms of integration across curriculum areas. They were more likely to feel they had enough guidance and advice to feel confident about National Standards (19% compared with 12% of

all principals), but not more likely to think that the timeframe for implementation was sufficient.

Small schools were less likely to have started implementing National Standards. They were less likely to report moderation against school benchmarks prior to the introduction of National Standards, and less likely to be moderating OTJs across year levels. However, they were more likely to be working with other schools to moderate OTJs.

Trustees in small schools were less likely to think that National Standards would improve achievement in their school, and less likely to agree the timeframe was realistic (4% compared with 17% of trustees in other schools, in both cases). However, they were also less likely to think that National Standards were difficult to put into practice (17% compared with 32% of trustees in other schools).

Year level

New entrant/Year 1 teachers were most likely to report that it was clear where all their students were in relation to the standard in reading (80%), and Year 7-8 teachers were least likely (55%). This was not the case for writing or mathematics. Year 7-8 teachers were less likely to be moderating OTJs and also less likely to report previous moderation against school benchmarks. New entrant/Year 1 teachers were less likely to say that National Standards had given them insight into students' learning needs (7%) than Year 2-8 teachers (15%).

Year 7-8 parents were more likely to have received a written report on their child's mid-year progress (87% compared with 48% of new entrant/Year 1 parents), and more likely to have had a discussion with their child and their child's teacher at a set time (75% compared with 38% of new entrant/Year 1 parents). They were also more likely to say that they received clear information on science progress, and on their child's achievement in relation to reading and writing National Standards.

3.2.2 Open responses on National Standards

Each of the stakeholder groups surveyed was asked if there were any comments they wanted to make on National Standards. Comments were made by 38% of parents, 51% of trustees, 45% of principals and 50% percent of teachers. Below is a summary of the thematic analysis of open response data presented in Wylie and Hodgen (2010).

Of the parents that commented on National Standards, 14% indicated that they had no knowledge or experience of the standards yet. Ten percent were confused about the standards or needed more information to be able to understand them. About a quarter made unqualified positive comments, while 10% made unqualified negative comments. Fifteen percent were concerned that National Standards work was at the expense of support for students, teaching or a broad curriculum. Seven percent were worried about negative effects on children, particularly if those children saw themselves as failing as a result of not meeting standards. Eleven percent wanted more specificity – feeling that the categories used in relation to National Standards were too broad. Some parents expressed dissatisfaction with the mid-year report, and some indicated that they wanted more information than they got from the report. Comments from parents indicated much variability in parents' experiences and expectations of National Standards.

Trustee comments were more often about the process of implementing standards, or how National Standards compared with what the school was doing previously. Of those trustees who made comments, 11% made unqualified positive comments, and 14% made unqualified negative comments. Fourteen percent expressed concern about negative consequences for students or schools. Sixty percent described some issue with the implementation or nature of National Standards. These issues were mostly related to the short timeframe for implementation, the lack of support for schools, queries about the standards themselves and uncertainty about what National Standards would provide, given that schools were already using assessment for learning and reporting.

Principal comments were mainly about implementation, and the standards themselves. Of those principals who made comments, 7% made positive comments. Thirty-nine percent made comments suggesting difficulty with the specifics of National Standards rather than the concept.

Principals commented on the speed of implementation, the need for more support for schools, and the fact that schools were already using assessment for learning and reporting. Some principals also suggested that raising achievement was more likely by other means than through the introduction of National Standards. Some indicated concern about consistency of OTJs across schools. Fifteen percent suggested that National Standards were introduced for political not educational reasons.

Of the teachers that commented on National Standards, 7% made positive comments. Forty-two percent were concerned at the lack of support for schools (particularly the lack of professional development and the short time-frame for implementation). Nineteen percent thought the school already had a good system of using assessment for learning and reporting prior to National Standards. Twenty-five percent had queries about the standards themselves. Twelve percent were concerned about possible negative consequences for students, and 12% commented that students don't all follow the same linear development path. Eight percent wondered why teachers' expertise was not utilised in the development of National Standards. Fifteen percent didn't think that National Standards would lead to raised achievement - rather, additional support would be needed. Some teachers also expressed concern that the National Standards would take away from teaching and other aspects of NZC.

3.2.3 Further exploration of National Standards and moderation

Following the initial examination of frequency patterns and cross-tabulations reported in Wylie and Hodgen (2010), Hipkins and Hodgen (2011) carried out a factor analysis, using the NZCER 2010 Primary and Intermediate Schools National Survey data, to explore relationships between teachers' experience of moderation prior to the introduction of National Standards and their views of National Standards at the time of the survey.

Hipkins and Hodgen (2011) found that previous experience with moderation of OTJs was associated with less difficulty in sourcing evidence for OTJs and greater clarity about their students' achievement against the standards. They also found that teachers with little or no experience of moderation prior to the introduction of National Standards were more likely to report changes in their assessment work as a result of National Standards, including doing more goal-setting with students. These teachers were more likely to see National Standards as having a positive impact. Teachers in schools where moderation practice was established before National Standards were introduced were less likely to say that they were working more with students on goal setting, and less likely to be positive about the impact of National Standards.

The overall conclusion drawn from the factor analysis was that the need for moderation of OTJs may provide opportunity for professional growth for teachers who had not previously experienced moderation practice. Hipkins and Hodgen (2011) suggest that this exposure to new and positive ways of working has the potential to result in positive impacts on achievement.

4 SECONDARY ANALYSIS

The report on the NZCER 2010 Primary and Intermediate Schools National Survey by Wylie and Hodgen (2010) provides a snapshot of the views and experiences of stakeholder groups in relation to the introduction of National Standards. Some of the findings suggest that there may be patterns in views and experiences worthy of further exploration in the interests of gaining a fuller understanding of the opportunities and risks involved in this significant policy shift. Hipkins and Hodgen (2011) looked at some of those patterns in their report on the practice of moderation and its relationship with National Standards. Their analysis pointed to some associations between teacher moderation experience and teacher views on National Standards, with the potential to usefully inform future professional learning and development discussions. With a secondary analysis of the survey data focusing on moderation yielding valuable insight, what else might be gleaned from secondary analyses of other aspects of the data? One aspect of the dataset that appeared to hold promise for further exploration was the open response data. All stakeholder groups were given the opportunity to comment on National Standards. The data contained within the open responses could convey experience of the implementation of National Standards in the respondent's own words, and on their own terms. If relationships between the themes and ideas expressed in open responses, and other variables within the dataset could be identified, it could provide an as yet untapped source of information with which to build understanding of the impacts of National Standards.

4.1 Coding by theme – Methods and results

To look for any associations that might exist between the ideas expressed in the open responses and other variables within the dataset, the open responses were first coded by theme with the intention of using statistical methods to explore possible relationships between those themes and the other variables. The decision was made to recode the open responses (rather than using the existing coding undertaken by NZCER), as this would allow for more interaction with the open response data, hopefully leading to a greater level of understanding and insight with which to enter into further analysis.

The coding framework was not intended to be mutually exclusive or exhaustive, rather – to indicate the prevalence of certain ideas or observations across the dataset that might tell a story about the experiences and attitudes of the various stakeholder groups. To this end, a list of some possible themes was created, and the responses were coded on the basis of whether

each theme was present or not. New coding categories were added as additional themes or ideas, not covered elsewhere, became apparent. The coding process was an iterative one that involved constantly revising coding categories to accommodate new data, and recoding responses to accommodate new coding categories. Once all responses had been considered, very similar coding categories were collapsed, and codes that were considered redundant (due to a very small number of relevant responses, or data being adequately captured elsewhere) were removed.

In the case of the principal questionnaire data, open responses from a question that asked for comment on how the introduction of National Standards had affected the schools' developmental work on the New Zealand Curriculum were coded along with responses to the question that asked for comment on National Standards. This was because, from looking through the data, it was apparent that many of the same themes were present among the responses. Principal responses for both questions were treated collectively. Each theme was only coded once for each respondent, and each respondent was counted once in determining the total number of responses for calculating percentages (see Table 1).

This initial coding process was undertaken at the level of stakeholder type (principal, teacher, trustee, and parent). All of the responses from principals were analysed and a set of coding categories developed, then all of the responses from teachers, and so on. The next step in the process was to compare the coding categories across the different stakeholder types, with the aim of being able to provide a commentary that considered the range of views between, as well as within, groups. A large number of coding categories had been identified for some stakeholder types, while a relatively small number had been identified for others. Some of the coding categories were distinct to a stakeholder type (not present in others), while some were common to multiple, or all, stakeholder types. In the interests of cross-group comparison, an attempt was made to maximise the cross-over in coding frameworks between the stakeholder types. This process involved directly comparing the coding frameworks for each stakeholder type, and on that basis, establishing a new, common one, with the necessary deletions and additions to accommodate as much of the data as possible. Some of the coding categories needed to be adapted in order to find a point of commonality (i.e., the stakeholder types had coding categories that were similar but could not be combined without making some slight alterations to the description, and as a result, to what was included). Then the responses for each survey type had to be checked against the new framework. During the checking of the

recoded data, some issues were discovered – responses didn't fit well into the coding category, or the categories had become too broad, encompassing a range of unrelated ideas.

A high degree of interpretation and assumption was required in identifying (and coding) themes, with responses rarely fitting neatly together in a defined category. There was a significant challenge in creating categories specific enough to be interesting and meaningful, yet broad enough to accommodate a volume of responses that would indicate some significance, or provide for further analysis. The nature of this coding framework (particularly its lack of exhaustiveness and mutual exclusivity) would make it very difficult for someone else to reproduce. For this reason, it was decided that testing inter-rater reliability would not be a useful exercise in this case. Instead, the data should be treated as subjective by nature.

On the basis of inherent subjectivity in coding, and bearing in mind the small number of responses in most of the coding categories, it was decided that statistical analysis of the theme-based coding data in relation to other variables in the dataset would not be practical or valid. Instead a discussion is presented here, of the findings of the theme-based coding, in the interests of providing a more detailed picture of the views on National Standards expressed in the NZCER 2010 Primary and Intermediate Schools National Survey. There were 32 coding categories created. Ten of those categories are outlined in section 4.1.1, in the interests of providing a general illustration of the nature of the coding process and the resulting data. The ten categories were chosen on the basis of the need for explanation or elaboration, prevalence, and the identification of patterns of response within and between stakeholder groups worthy of further discussion. Overall impressions of the data from the coding process are then presented in section 4.1.2, with a full list of all coding categories showing the number of relevant responses from each stakeholder group, and the proportion of open responses represented by that number (Table 2).

4.1.1 Coding categories

(1) Rushed or poor implementation, lack of trialling, lack of evidence base, or not taking into account examples from overseas

This coding category encompasses issues relating to the quality of National Standards planning, development, and implementation processes. Responses were generally coded against this category because they clearly identified one of the issues specified in the category

descriptor. There was some consideration of how valid it was to combine these different issues in one category. In the case of the teacher dataset, for instance, it was possible to separate lack of trialling, rushed implementation, and lack of evidence base and still have relatively large numbers in each category. However, when it came to the other datasets, the numbers were very small for the individual codes and so it seemed to make sense to combine them.

Comments that suggested a lack of preparatory work prior to implementation – that systems, processes, resources, and information were not in place before National Standards were rolled out – featured in this category, as well as more generic references to inadequate or poor implementation. A number of responses indicated that the standards were not well thought through. Where these comments specified that it was the implementation that was poorly thought through, or that standards were implemented without having been properly thought through first, they were included in this category.

For all stakeholder types except parents, reference to poor or rushed implementation was the most common reason for classification under this category, followed by lack of trialling. Relatively speaking, lack of trialling was mentioned more often by teachers than principals or trustees. Often poor or rushed implementation was linked with a lack of support. This is reflected in some crossover between this coding category and "Lack of clear and consistent information and support (including professional development) for the sector around National Standards".

(2) Issues with how and where the standards have been set, including levels (incorrect or ambiguous) and processes used

This category relates to the design of the standards themselves – how valid they are, and how "fit for purpose" as a set of national benchmarks. Responses included in this category related specifically to the practical application of the standards. For instance, there were comments (across the different stakeholder types) indicating that the standards were inappropriately high or low (either in general, or with specific reference to a curriculum area and year level). There were also comments suggesting that the aspirational nature of the standards was an undesirable feature. Related to observations about the usefulness of the levels, were comments on the appropriateness of the processes used to create the benchmarks. Some respondents noted that they felt the standards and/or the processes used to set them were arbitrary, and some indicated that they felt the standards should have been based on

achievement data or norms. Some noted that the standards were too vague, broad, or general to provide useful information.

Many parents expressed a desire for more specific information about their child's achievement – for instance, if they are "below", then how far below. Though this type of comment is similar to observations that the standards are too broad or general, it has not been included in this category as it is considered a reporting issue rather than a design issue. Also not included in this category were responses that related to the philosophical underpinnings of the National Standards policy (*why* standards have been set). These responses fitted better in other categories.

Other responses that were not as common but did come into this category included the observation that having a "well below" but no "well above" was unbalanced and the assertion that this would be detrimental to high achievers; and, a suggestion that learning in literacy and numeracy is not linear and step by step as the standards would suggest. This latter comment was similar to a common sentiment that children's learning patterns and styles are diverse and not well served by a "one size fits all" model such as the National Standards, which was not included in this category as it was considered to be about diversity of students, and as such, more about questioning the validity of having standards at all than questioning where and how those standards have been set.

Eleven teachers and one principal noted issues with the assessment timeframes for National Standards. Assessment after Years 1, 2, and 3 at school was seen as difficult because it was not in line with the school's assessment or reporting cycles (where an issue with reporting was noted, this was also coded under the category "Issues or concerns with aspects of National Standards reporting (being flawed, inadequate, or inferior to school's previous reporting)"), or because it was impractical or confusing. Assessment after three years at school, and then at the end of Year 4 was raised as problematic given that for some students it is a relatively short time between these assessments and others a relatively long time, leading to huge variation in what students are expected to achieve. Further to this, several teachers noted that end of year assessment and reporting (required from Year 4 on) could mean students were up to a year different in age, but with the same expectations against National Standards.

(3) Issues or concerns with aspects of National Standards reporting (being flawed, inadequate, or inferior to school's previous reporting)

This coding category has as its central theme issues arising in the conveying of National Standards information to parents through reporting. Comments coded against this category were most common among parents, which might be expected given that most parents' primary interaction with, or experience of, National Standards is through reporting.

A number of responses from teachers (and one from a trustee) that were included in this category expressed concern or frustration over the difficulties involved in having to report against an end-of-year standard at mid-year, while teachers and principals questioned the practicalities of having to report after Years 1, 2, and 3 and then at the end of Years 4, 5, 6, 7, and 8. One principal mentioned incompatibility of formative assessment and reporting timetables, suggesting that additional assessment was having to be undertaken purely for reporting purposes. Parents who raised issues with reporting timelines made comments about how unhelpful it was that mid-year reporting was based on tests done at the beginning of the year and therefore told them nothing current or new about their child's achievement or progress.

Another common thread running through this category was a feeling that National Standards reporting lacked clarity. While a number of parents indicated that they found the reporting either relatively meaningless, or confusing and difficult to understand, a number of teachers noted difficulty conveying information to parents in a way that was useful or easy to understand. One principal and several trustees similarly expressed concerns that reporting was confusing for parents.

A desire for something more or different in reporting was expressed by respondents from all stakeholder groups. Some respondents explicitly noted that they felt the school's previous reporting methods were superior, while others suggested that they would prefer to report on/would prefer to hear about other forms of data – for instance, standardised test data, curriculum levels or grades. Some parents noted that they liked National Standards reporting, or at least some aspects of it, but would like to have something else as well. Additional information wanted by parents included grades, percentages, and rankings; an indication of where within the National Standards category the student was (i.e., how far below or how far above); contextual information about the standards and how to interpret them; and information about other areas of curriculum or development.

(4) Concerns over the impacts of labelling on students (particularly in relation to self-esteem and self-efficacy)

The two main ideas encompassed within this category are that National Standards result in the labelling of students and that being assessed as below standard could have detrimental effects on the self-esteem of students as well as their motivation and attitude to learning (with the importance of self-belief and positive attitude to learning being stressed by teachers and principals).

The first idea was sometimes expressed as a stand-alone concept (for instance, a number of comments indicated that National Standards label students without being explicit as to why or how this labelling is of concern), and sometimes it was expressed in relation to a specific negative outcome or set of outcomes. In most cases, the negative outcome/s related to the second idea – being the negative effect on self-esteem, motivation, and attitude. Many (but not all) of those who mentioned labelling used the context of "failure" – i.e., students will be labelled as failures. Some respondents also talked about labelling students as failures so early in their school careers, either specifying or implying that this would have an effect on their subsequent experiences at, and of, school. There is some crossover here with another category – "National Standards don't accommodate or recognise diversity in learners or contexts", with some respondents noting that the labelling of students as failures is unfair as it doesn't acknowledge that learners are diverse, achieve at different rates and different levels, and face different challenges.

National Standards were considered by some respondents to facilitate a focus on students' deficits rather than strengths (for low achievers – the focus on, and perceived importance of, the fact that they are below standard rather than the progress they have made and the skills they have in areas other than literacy and numeracy). Some respondents also mentioned negative effects of labelling on parents and families as well as students.

(5) The concept of National Standards (or elements of it) is good but have reservations

Comments in this category indicated support for the concept or principle of National Standards, but with reservations, concern, or dissatisfaction with some specific aspect of the standards or their implementation. Rushed or poor implementation was the most commonly mentioned reservation among trustees and teachers, and was also mentioned by two out of the three principals whose responses were coded against this category. All three of the principals

that expressed support for the concept (or elements of the concept) of National Standards also mentioned that National Standards were politically motivated. The most common concern among parent respondents in this category was a lack of information or understanding about National Standards.

Other issues mentioned included concern over non-formative use of the National Standards data (for instance, the creation of league tables), lack of consultation or buy-in from the sector, lack of clear and consistent information and support around the standards, the potential for inconsistency and subjectivity in OTJs, and the inability of the standards to adequately accommodate diversity.

(6) Lack of clear and consistent information and support (including professional development) for the sector around National Standards

Comments coded against this category generally conveyed the idea that the advice, support and resourcing needed by schools to effectively implement the National Standards was not provided in a timely manner or, in some cases, at all. Where guidance was available, it was sometimes unclear or inconsistent, involving "mixed messages" and "moving goalposts". A number of respondents noted that they were having to use the standards (in assessing and reporting) without having received sufficient training, or without having access to required resources and information.

Comments from principals, teachers, and trustees suggesting a frustration with a lack of quality professional development on the National Standards were featured within this category. Comments related to professional development included observations that trainers were ill-prepared and lacked answers to important questions. It was also noted that information and materials required by schools were not available at the time of training. Dissatisfaction was expressed by several teachers with the fact that professional development was not available to all teachers. Related comments pointed to the high cost of the available professional development and the limited resources of schools, and the unreasonableness of expecting senior management within schools to be responsible for training teachers on the use of National Standards.

A lack of timely practical advice and support (for instance, templates and information on how the National Standards align with other assessments) was also mentioned more generally (not specifically in relation to formal professional development) by some principal and teacher

respondents. A related observation from principals, teachers, and trustees was that the advice and information about the National Standards that was provided to schools by the Ministry of Education and advisory services involved changes of position that resulted in confusion in schools.

This category also contained comments from trustees who felt that there was insufficient training and information made available to Boards of Trustees. In addition, several trustees noted that schools were having to educate parents about the National Standards, where this role should have been fulfilled by the Ministry of Education.

This category was associated with a perception of rushed or poor implementation, with lack of clear and consistent information and support presented as an element of poor implementation, or as a result of the rush to implement the National Standards policy.

There were no parent responses coded against this category.

(7) Inconsistency, subjectivity, or moderation issues in OTJs

Responses in this category were most common among teachers and least common among parents. They generally revolved around the idea that OTJs would not, or could not, provide a reliable means of measuring (and comparing) student achievement.

Respondents across all stakeholder types pointed to the subjective nature of OTJs as being potentially problematic. Some comments indicated that this would lead to inconsistencies between teachers, while others saw problems arising from inconsistency between schools, and resulting issues of fairness if National Standards data were used to compare schools with each other. Comments from teachers, trustees, and parents expressed concern over the possibility that a desire to achieve good National Standards outcomes may result in "teaching to the test", or that data may even be manipulated in order to portray achievement levels in a more favourable light.

Some comments from principals, teachers, and trustees suggested that the lack of clarity and specificity in the standards themselves, or in the guidance provided to schools in interpreting the standards, meant that achieving consistency was not possible. A number of responses from members of these stakeholder groups also made specific reference to moderation – moderation was seen as crucial for consistency, both within and between schools. However, it

was also seen as difficult, impractical, time-consuming and inadequately resourced, particularly at an inter-school and inter-regional level.

Another observation made by a small number of principals and teachers that features in this category is that the only way to achieve consistency in assessing against National Standards is to use a system of national testing. While some of these respondents indicated that this was an undesirable but inevitable outcome, some of them voiced a preference for national testing over the use of OTJs, on the basis that it would be more reliable and less time-consuming.

(8) Parents, or the public in general, are confused by, don't understand, or lack knowledge of National Standards

This coding category encompasses a variety of responses that have been grouped together because they all relate to a lack of knowledge and understanding about National Standards among parents. In some cases, comments are generalised to the wider public. While parents are the focus of this category and most of the comments coded against it, comments that refer to the community or to the general public have been included on the basis that the community and the general public are stakeholders in education who are likely to have access to many of the same messages and much of the same information as parents (being information conveyed through the media).

Respondents from across the range of stakeholder types made comments indicating that parents (as a group) have insufficient knowledge of National Standards. In some cases it was suggested that this was because information had not been communicated by the Ministry of Education, and instead educating parents had been left up to schools who have not, themselves, had access to clear and consistent information. A number of parents commented that they had received little or no information about National Standards, and some expressed a desire for more information. Those responses are coded against this category.

A lack of understanding of, or confusion about, the National Standards among parents was also noted by respondents from all stakeholder groups. In some cases, similar comments were generalised to the wider public. One trustee suggested that the Minister of Education did not understand the National Standards. There were several comments from principals and teachers that indicated parents, or the public, had been deliberately misled by the government as to the nature of the National Standards, and what they might do for the education system.

One teacher noted that parents were not aware of the negative implications of the National Standards.

Some specific misconceptions or areas of confusion for parents were mentioned by respondents. The aspirational nature of the National Standards was one of these areas, with principals and teachers suggesting that parents believed the National Standards represent average achievement levels. A similar confusion was indicated by a comment from a principal that parents do not understand the difference between criterion-referenced and norm-referenced assessment. Teachers, and parents themselves noted confusion over the lack of alignment between the National Standards and other forms of assessment, with some parents struggling to comprehend how their child could be average according to an e-asTTle, PAT, or Observation Survey test (commonly used norm-referenced assessments), but still be below standard, for instance. Principals, teachers, trustees, and parents described further issues arising from National Standards reporting timetables. Mid-year reporting against an end-of-year standard was considered difficult for parents to grasp, as was the change in reporting timetables between the end of the student's third year at school and the end of Year 4.

Comments from teachers and trustees on the consequences of parental confusion over the standards were captured within this coding category, and included perceptions of anxiety and anger on the part of parents who saw their children as failing.

(9) National Standards have improved or will/may improve some aspect of school practice

This category includes specific comments about how a school's (or schools') practice has or might be improved as a result of National Standards. An increase in professional discussion was the central feature of principal comments coded against this category. Such discussion was seen as useful for self-review and reflection, developing shared understandings of progress and achievement among staff, and enhancing understanding of the New Zealand Curriculum. Other benefits perceived by principals included improvements in moderation practice and reporting. Most of the principal responses in this category, while communicating some actual or potential benefit, also communicated reservations about other issues or aspects of the standards – the nature of these reservations was mixed, but some crossover was apparent with concerns over non-formative use of data, as well as with observations on the politically motivated nature of the standards.

Some teachers also saw discussion and reflection stimulated by National Standards implementation as beneficial for professional growth and development. Other comments from teachers coded against this category involved perceived or anticipated improvement in consistency of expectation and judgement in relation to student achievement. Increased and improved moderation practice was mentioned, as was a greater degree of partnership with parents. While fewer teachers than principals qualified potential benefits with reservations, there was some crossover with poor or rushed implementation, and lack of information and support for the sector.

All the responses from trustees that were coded against this category were unreservedly positive. Trustees saw actual or potential improvements in reporting to parents, and also in the information available to Boards of Trustees for use in strategic planning. One trustee suggested that the changes made in implementing National Standards had added to the professionalism of the school. Another noted that the standards provided an opportunity for benchmarking of student achievement in rural schools.

The majority of parent responses coded against this category pointed to some benefit or improvement in reporting practice. Features of reporting seen as providing benefit included the specification of goals or next steps, and the relaying of information and ideas for parents to use in assisting their child's learning at home. Three of the parent responses in this category suggested that National Standards could help prevent negative outcomes (with two describing National Standards as a way to stop children “falling through the cracks”, and one suggesting that the standards might prevent her child from “being a 15 year old who doesn’t know their basic facts”).

Both of the parents who saw National Standards as a measure to stop students falling through the cracks also reported that they had received little or no information about National Standards.

(10) National Standards don't accommodate or recognise diversity in learners or contexts

Comments coded against this category questioned the usefulness of the National Standards as a measure of success, specifically in relation to their ability to adequately reflect achievement of diverse groups and individuals.

Comments in this category mentioned a range of factors that impact on achievement but are not necessarily within the sphere of influence of schools. Concern was expressed that having larger-than-average proportions of English language learners, students with special educational needs, or students from low socio-economic backgrounds would disadvantage schools if there were comparisons on the basis of National Standards results. Several comments from teachers indicated a concern that if remuneration were linked to National Standards results, teachers would avoid schools or classes where the social, cultural, and economic contexts might lend themselves toward lower levels of achievement against the National Standards.

Along with concerns about the fairness of comparing schools, concerns were expressed about the fairness of comparing students. The fact that the National Standards require the same expectation of all students on the basis of their age, and irrespective of any context, was seen as problematic in its inability to accommodate the learning needs of a range of students, some of whom may have challenging circumstances affecting their ability to reach the standards. Some of the comments from principals, teachers, and parents made specific reference to the fact that children start school with very different levels of knowledge and skill (depending on their home environment, and whether they have been in early childhood education for instance), making it unreasonable or unrealistic to expect them all to be achieving at the same level one year later.

Another observation coded against this category was that, while the standards assume a step-by-step linear learning progression common to all students, in reality students learn at different rates and in different ways, and the standards are not able to accommodate this. Again, it was noted that the National Standards are particularly inappropriate for the first years of school, as there is an even greater degree of variation in rates and patterns of progress among students during this time. This idea was taken a step further by a handful of parents and teachers who suggested that not all children are of equal ability, and therefore, to expect them all to achieve the same outcomes is unrealistic. Two teachers made reference to a normal distribution of ability. Several teachers noted that, where children are not high achievers, it is important to acknowledge the progress they do make, and one teacher suggested that it is preferable to set realistic and attainable goals based on each student's individual learning needs, than to focus on where all students “should” be.

Comments about the narrow achievement focus of the National Standards were coded against this category when those comments made express reference to the impact on students. Such comments were made by principals, teachers, trustees, and parents; and questioned the limiting definition of success inherent in the standards, including the inability of the standards to acknowledge things like attitude, work habits, social skills, and creativity. The standards were seen to devalue these factors (and thereby, students with strengths in these areas), along with other skills, talents, and interests outside of literacy and numeracy.

The need for additional resources to support students for whom meeting standards is, or will be, challenging was raised by one principal, two teachers and a parent. One teacher and one parent felt that the National Standards would result in a focus on students close to the "boundary" (between "below" and "at" the Standard), and a neglect of the students who really struggle or clearly excel.

4.1.2 Overall impressions of the data

At the time the survey was carried out it would appear that, as far as schools were concerned (principals, teachers, and trustees), the most problematic aspect of National Standards was the quality and pace of implementation, including a lack of access to the resources required to understand and use the standards. There were a number of respondents (particularly trustees) who made comments that explicitly or implicitly communicated an openness to the concept and aims of National Standards, but dissatisfaction with the practical implementation of the policy.

Another relatively common sentiment from within the sector was that the standards have inherent flaws that limit their ability to provide a useful measure of success in education, both at the school and the student level. These flaws included the restrictive definition of success supported by the National Standards (specifically in relation to their narrow academic focus), and the coarse nature of the levels of achievement that do not communicate the progress a student might be making. Teachers, in particular, were concerned about the levels the standards had been set to and the process used to establish these levels, which was seen by some as fairly arbitrary, and in some cases the resulting standards were considered unrealistic. A number of respondents from across all the stakeholder types questioned the fundamental assumption underpinning the standards, that we should have common expectations for all students. While some pointed to student-based factors that influence a student's performance in literacy and numeracy (such as variation in learning styles and

learning rates, having English as a second language, or having special educational needs), others pointed out that by expecting all students in all schools to reach the same levels at the same time, National Standards neglect to acknowledge the reality of social and economic inequalities and their impact on educational outcomes. The idea that it is unfair to judge all students in the same way without taking circumstances and context into account was linked with concern that students labelled as failures through not meeting standards could suffer unnecessary harm as a result, and could have their future academic progress impeded through damage to motivation and self-efficacy.

While the role of teacher judgement in National Standards is an aspect of the policy that has been highlighted as having the potential to avoid many of the pitfalls of national testing, this was considered by a number of respondents to represent another design flaw, given the perceived subjective nature of these judgements and their ability to be manipulated in order to give an inflated impression of performance. Others noted that teacher judgements needed to be supported by good inter- and intra-school moderation practices which were considered challenging to put in place due to limited resourcing and other practicalities. Whether or not teacher judgement is preferable to a system of national testing, some of the responses from principals, teachers, and trustees convey the idea that OTJs are redundant as they add nothing new to the educational toolkit. Responses indicated frustration that, as schools were already measuring student achievement, tracking progress, and reporting to parents, the standards represented little more than extra workload for no real gain.

The pattern of responses among parents was quite different from that of the other three stakeholder types, which is logical given that parents interact with the standards in a different way. It is, however, important because it seems that there is some disjuncture in expectations between those in schools, and external stakeholders. The most pressing issue for parents would seem to be a lack of information or understanding of the National Standards. Many parents acknowledged that they were confused by the standards, or that they needed more information to feel confident engaging with them. In a number of cases, comments from parents inadvertently reflected a lack of understanding of the nature of National Standards. For instance, a number of parents appeared to confuse “at” with “average”. While parents seemed more positive overall than the other stakeholder groups about what National Standards might deliver, many of them appeared to want more from National Standards reporting than they had yet received. The sort of additional information that parent respondents expressed a desire for further supports the idea of fairly widespread

misconceptions among parents about what National Standards are able to provide. Quite a number of respondents wanted more exact information about their child's achievement level – a grade, or a percentage for instance. Some of these parents wanted to know specifically how their child ranked compared with others in relation to National Standards. This is the sort of information that might be obtained through norm-referenced assessment, rather than criterion-referenced assessment like the National Standards, and lends credence to the claim by one principal respondent, that the public (parents being a subset of that group) lack understanding of the characteristics of different forms of assessment.

For reference, the total numbers of open responses by stakeholder type are given in Table 1.

Table 1: Number of open responses by stakeholder type

Stakeholder type	Total number of open responses
Principal	126
Teacher	481
Trustee	130
Parent	211

In Table 2, a full list of coding categories is presented, showing the number (n) and percentage (%) of responses coded against that category by stakeholder type. Percentage is calculated as the number of open responses from the relevant stakeholder type coded against the relevant category, out of the total number of responses from that stakeholder type. Percentages are rounded to 1 decimal place.

Table 2: Full list of theme-based coding categories

	Coding category	Principals		Teachers		Trustees		Parents	
		n	%	n	%	n	%	n	%
	Lack of consultation and sector input in development, or general lack of buy-in and support from the sector.	13	**10.3**	48	**10.0**	15	**11.5**	6	**2.8**
(1)	Rushed or poor implementation including lack of trialling, lack of evidence base, or not taking into account examples from overseas.	36	**28.6**	168	**34.9**	40	**30.8**	10	**4.7**
(2)	Issues with how and where the standards have been set, including levels (incorrect or ambiguous) and processes used.	12	**9.5**	84	**17.5**	8	**6.2**	11	**5.2**
(3)	Issues or concerns with aspects of National Standards reporting (being flawed, inadequate, or inferior to school's previous reporting).	6	**4.8**	24	**5.0**	8	**6.2**	38	**18.0**
	Extra resourcing needed to improve achievement (including for those identified as “below” or “well below” under National Standards).	7	**5.6**	37	**7.7**	9	**6.9**	8	**3.8**
(4)	Concerns over the impacts of labelling on students (particularly in relation to self-esteem and self-efficacy).	8	**6.3**	58	**12.1**	5	**3.8**	11	**5.2**
(5)	The concept of National Standards (or elements of it) is good but have reservations.	3	**2.4**	31	**6.4**	21	**16.2**	11	**5.2**
	Concerns over non-formative use of data by the government or public (i.e., league tables), and the impact this may have on schools.	9	**7.1**	42	**8.7**	7	**5.4**	5	**2.4**
	National Standards provide useful information for parents, or parents like them.	1	**0.8**	9	**1.9**	12	**9.2**	38	**18.0**
(6)	Lack of clear and consistent information and support (including professional development) for the sector around National Standards.	26	**20.6**	103	**21.4**	20	**15.4**	0	**0.0**

	Coding category	Principals		Teachers		Trustees		Parents	
		n	%	n	%	n	%	n	%
	The sector needs more time, information, or support (including professional development) to understand and use National Standards.	8	**6.3**	35	**7.3**	7	**5.4**	1	**0.5**
	National Standards are time consuming and mean an increased workload for school staff or trustees.	15	**11.9**	56	**11.6**	11	**8.5**	8	**3.8**
	National Standards will not improve student achievement, or it is not clear how or why they would.	12	**9.5**	41	**8.5**	4	**3.1**	0	**0.0**
	National Standards are a poor use of resources.	3	**2.4**	34	**7.1**	6	**4.6**	9	**4.3**
	There is too much focus on measuring and reporting, instead of teaching and learning.	2	**1.6**	22	**4.6**	6	**4.6**	10	**4.7**
	National Standards are politically motivated, or National Standards involve or encourage school or teacher bashing.	13	**10.3**	27	**5.6**	4	**3.1**	1	**0.5**
	National Standards encourage narrowing of curriculum focus, or have a detrimental effect on the New Zealand Curriculum.	9	**7.1**	36	**7.5**	3	**2.3**	7	**3.3**
	There is a lack of alignment between the National Standards and other key resources (including the New Zealand Curriculum).	8	**6.3**	29	**6.0**	3	**2.3**	1	**0.5**
(7)	Inconsistency, subjectivity, or moderation issues in OTJs.	16	**12.7**	67	**13.9**	10	**7.7**	3	**1.4**
	National Standards don't recognise or show students' progress.	4	**3.2**	36	**7.5**	2	**1.5**	4	**1.9**
	There shouldn't be too much emphasis placed on the standards.	9	**7.1**	4	**0.8**	2	**1.5**	6	**2.8**

	Coding category	Principals		Teachers		Trustees		Parents	
		n	%	n	%	n	%	n	%
	Having a national or consistent benchmark is good.	1	**0.8**	7	**1.5**	6	**4.6**	11	**5.2**
	The school, or schools in general, already measure achievement, or know where students are at.	13	**10.3**	83	**17.3**	18	**13.8**	7	**3.3**
	The school, or schools in general, already report clearly to parents.	8	**6.3**	34	**7.1**	9	**6.9**	9	**4.3**
	Teachers, principals, or trustees are confused by or don't understand National Standards.	5	**4.0**	25	**5.2**	5	**3.8**	1	**0.5**
(8)	Parents, or the public in general, are confused by, don't understand, or lack knowledge of National Standards.	6	**4.8**	33	**6.9**	5	**3.8**	51	**24.2**
	National Standards create pressure or anxiety for stakeholder groups (including teachers, principals, BOT, students, and parents).	5	**4.0**	29	**6.0**	5	**3.8**	2	**0.9**
	National Standards have made, or will make, little difference to practice or experience.	6	**4.8**	18	**3.7**	6	**4.6**	5	**2.4**
(9)	National Standards have improved or will/may improve some aspect of school practice.	14	**11.1**	10	**2.1**	11	**8.5**	9	**4.3**
	National Standards are not a useful measure.	7	**5.6**	19	**4.0**	2	**1.5**	20	**9.5**
(10)	National Standards don't accommodate or recognise diversity in learners or contexts.	18	**14.3**	95	**19.8**	12	**9.2**	18	**8.5**
	There is an unhelpful negative attitude to National Standards in the sector.	2	**1.6**	3	**0.6**	3	**2.3**	4	**1.9**

4.2 Creating an attitude measure from open response data

Having found that coding the open responses on National Standards from the NZCER 2010 Primary and Intermediate Schools National Survey by theme resulted in a large number of codes (each having a small number of related responses), and entailed such a degree of subjective interpretation as to make any statistical analysis of highly questionable worth, alternative ways of quantifying the open response data were considered.

Many of the responses seemed to indicate quite clearly whether the respondent felt the National Standards were a positive move for the education system, or a negative one. It was decided that associations between open responses and other variables within the data set could potentially be usefully explored by treating the open responses as a binary response variable, being negative or positive attitude (to the National Standards), and the other variables within the data set considered as possible predictors using measures of goodness of fit, including logistic regression for continuous and ordinal explanatory variables, and Fisher's exact tests for categorical explanatory variables.

The intention was that this coding framework would be exhaustive and mutually exclusive, with responses lacking any discernible value judgement allocated to a third category of "none", and excluded from analysis. However, the responses did not always fall neatly into one of the categories. There were a number of responses within each set of questionnaires that contained strong views, but ones that were not consistently positive, or negative. Of particular note were the conditional responses – for instance, where the respondent indicated that they liked the idea of having National Standards, but felt that the implementation process had been flawed. It did not seem useful to discount these responses (the effect of allocating them to "none"), but neither did it seem valid to assign them to a purely positive or negative category, given the bias inherent in choosing a dominant theme. Therefore, a fourth category was introduced. This category was labelled as "both/conditional".

In a number of cases, it was difficult to discern between an attitudinal position, and an observation without any underlying value judgement. Here, the language and phrasing were largely relied upon in determining tone, and thereby deciding whether a response could be usefully assigned to an attitudinal category. For instance (in response to a question about the impact of National Standards on the implementation of the New Zealand Curriculum), "we have shifted our focus" has been treated as neutral ("none"). However, "…has forced us to 're-direct' our strategic goals…" has been treated as negative, due to the use of the word

"forced" which suggests being compelled to comply. Even though the observation is essentially the same, the attitude appears (on the surface) different. There is obviously a high degree of subjectivity involved in this interpretation. However, as there is no objective measure of the attitude underlying each response, it was considered preferable to demarcate in this way, than to exclude all such responses from analysis.

4.2.1 Reliability

Given the subjective elements of the coding process, and the potential effects of bias on findings, it is considered good practice to undertake some testing of reliability (Lombard et al., 2004). Measuring the level of agreement among multiple coders is a means of testing the reliability of the data (Hayes & Krippendorff, 2007). Although such testing will not remove the subjective elements from the coding process, it can provide an indication of whether a coder's interpretations are predictable, through determining how similarly another coder would interpret the same data.

In order to test the coding scheme for reproducibility, and thereby provide some information about the reliability of the data, it was decided that the entire dataset should be recoded by a second coder. The second coder was given the coding categories together with a brief explanation of the approach taken, similar to the explanation provided in the first few paragraphs of this section.

A number of alternative statistics for calculating inter-rater reliability are available in the coding literature (Potter & Levine-Donnerstein, 1999; Hayes & Krippendorff, 2007; Lombard et al., 2004; Multon, 2010). The most appropriate measure will often be dependent on contextual factors such as the number of coders; and whether the coding structure is nominal, ordinal, or scale-based. Proportion of overall agreement (P_o) is the simplest and least robust measure of agreement. It can give an indication of consistency, but this kind of measurement is limited in that it does not account for chance agreement. Three common methods of correcting for chance agreement are Scott's Pi, Cohen's Kappa, and Krippendorff's Alpha (Potter & Levine-Donnerstein, 1999; Hayes & Krippendorff, 2007; Uebersax, 2010).

Scott's Pi and Cohen's Kappa have both been criticised for over correcting for chance agreement. Scott's Pi formula treats deviation from equal selection of coding values as agreement by chance (Potter & Levine-Donnerstein, 1999). Given the nature of the data set, this would present an issue, as there would rarely be a high level of balance in coding

selection. Unlike Scott's Pi, Cohen's Kappa can accommodate more than two coders. However, Cohen's method has been subject to criticism on the basis that it provides for a linear reduction in chance agreement as the number of coders increases, where this reduction actually follows a geometric pattern, decreasing the usefulness of the Kappa statistic with an increase in coder numbers (Potter & Levine-Donnerstein, 1999). As there were only two coders in this case, issues arising from having a large numbers of coders were not relevant. However, the problem of over correction due to coding imbalance remained.

Krippendorff's Alpha is well regarded as a measure of reliability that has come out of content analysis (Lombard et al., 2004). It is considered to be flexible, and to avoid some of the limitations of Scott's Pi and Cohen's Kappa (Hayes & Krippendorff, 2007) However, it is still a conservative measure of agreement. Alpha is somewhat more complex to compute than other measures but where there are two coders and data is nominal, Alpha can be obtained more simply through making a slight adjustment to Scott's Pi. As the sample size tends toward infinity, Pi becomes a better approximation of Alpha, with the two statistics being equal in the case of an infinite sample. Where the sample is finite:

$$nominal\ \alpha = 1 - \frac{(n-1)}{n}(1-\pi) \geq \pi$$

(Krippendorff, 2004).

For each of the sets of responses (one for the teacher, trustee, and parent questionnaires; and two for the principal questionnaire) inter-rater reliability was calculated using proportion of overall agreement (P_o), Scott's Pi (π), and Krippendorff's Alpha (α) obtained using the adjustment to Pi described above.

The following tables show agreement. The relevant statistics (rounded to 2 decimal places) are presented below each table.

Teacher Questionnaire

		Coder 1			
		Positive	Negative	Both	None
Coder 2	Positive	9	0	1	3
	Negative	0	346	7	16
	Both	4	16	30	6
	None	0	5	0	38

$P_o = 0.88$

$\pi = 0.69$

$\alpha = 0.69$

Trustee Questionnaire

		Coder 1			
		Positive	Negative	Both	None
Coder 2	Positive	11	0	1	2
	Negative	0	64	2	5
	Both	0	1	24	4
	None	0	1	1	14

$P_o = 0.87$

$\pi = 0.80$

$\alpha = 0.80$

Parent Questionnaire

		Coder 1			
		Positive	Negative	Both	None
Coder 2	Positive	35	0	2	4
	Negative	1	75	0	18
	Both	7	2	10	15
	None	1	3	0	38

$P_o = 0.75$

$\pi = 0.64$

$\alpha = 0.64$

Principal Questionnaire 1 (Impact of National Standards on implementation of NZC)

		Coder 1			
		Positive	Negative	Both	None
Coder 2	Positive	2	0	0	0
	Negative	0	61	7	4
	Both	1	1	10	0
	None	0	0	0	8

$P_o = 0.86$

$\pi = 0.70$

$\alpha = 0.70$

Principal Questionnaire 2 (Comments on National Standards)

		Coder 1			
		Positive	Negative	Both	None
Coder 2	Positive	4	0	0	0
	Negative	0	58	4	4
	Both	0	1	2	1
	None	0	2	0	18

$P_o = 0.87$

$\pi = 0.74$

$\alpha = 0.74$

The tables and calculations above include responses classified as “none”. Krippendorff’s Alpha is calculated using pairable values. As the responses put in the “none” category have not actually been coded, they do not qualify as pairable values (Krippendorff, 2011). For this reason, the three statistics were recalculated, this time excluding the responses classified as “none”. The results are presented below.

Teacher Questionnaire

		Coder 1		
		Positive	Negative	Both
Coder 2	Positive	9	0	1
	Negative	0	346	7
	Both	4	16	30

$P_o = 0.93$

$\pi = 0.72$

$\alpha = 0.72$

Trustee Questionnaire

		Coder 1		
		Positive	Negative	Both
Coder 2	Positive	11	0	1
	Negative	0	64	2
	Both	0	1	24

$P_o = 0.96$

$\pi = 0.93$

$\alpha = 0.93$

Parent Questionnaire

		Coder 1		
		Positive	Negative	Both
Coder 2	Positive	35	0	2
	Negative	1	75	0
	Both	7	2	10

$P_o = 0.90$

$\pi = 0.84$

$\alpha = 0.84$

Principal Questionnaire 1 (Impact of National Standards on implementation of NZC)

		Coder 1		
		Positive	Negative	Both
Coder 2	Positive	2	0	0
	Negative	0	61	7
	Both	1	1	10

$P_o = 0.89$

$\pi = 0.68$

$\alpha = 0.68$

Principal Questionnaire 2 (Comments on National Standards)

		Coder 1		
		Positive	Negative	Both
Coder 2	Positive	4	0	0
	Negative	0	58	4
	Both	0	1	2

$P_o = 0.93$

$\pi = 0.68$

$\alpha = 0.68$

With "none" excluded, the level of reliability has increased for teacher, trustee, and parent surveys, but has decreased slightly for the two principal datasets.

To explore possible relationships between the open responses and characteristics or factors contained within the other survey responses, some sort of comparison needs to be made. It would be intuitive to compare those responses coded as positive to those with responses coded as negative and see if there are identifiable differences between the two groups of respondents that could form the basis of a model for predicting attitudes. However, the very small number of responses coded as positive make this sort of comparison difficult. To aid a more useful comparison, the coding categories were collapsed to a binary set of "negative" and "some positive" (a catch-all category for instances in which anything positive was

expressed, even if there was also some negative sentiment or observation). In this case, responses coded as "positive" and those coded as "both" are encompassed in the second category. The results of this collapse, including reliability statistics, follow.

Teacher Questionnaire

		Coder 1	
		Some Positive	Negative
Coder 2	Some Positive	44	16
	Negative	7	346

$P_o = 0.94$

$\pi = 0.76$

$\alpha = 0.76$

Trustee Questionnaire

		Coder 1	
		Some Positive	Negative
Coder 2	Some Positive	36	1
	Negative	2	64

$P_o = 0.97$

$\pi = 0.94$

$\alpha = 0.94$

Parent Questionnaire

		Coder 1	
		Some Positive	Negative
Coder 2	Some Positive	54	2
	Negative	1	75

$P_o = 0.98$

$\pi = 0.95$

$\alpha = 0.95$

Principal Questionnaire 1 (Impact of National Standards on implementation of NZC)

		Coder 1	
		Some Positive	Negative
Coder 2	Some Positive	13	1
	Negative	7	61

$P_o = 0.90$

$\pi = 0.70$

$\alpha = 0.70$

Principal Questionnaire 2 (Comments on National Standards)

		Coder 1	
		Some Positive	Negative
Coder 2	Some Positive	6	1
	Negative	4	58

$P_o = 0.93$

$\pi = 0.66$

$\alpha = 0.67$

Reliability is higher for all sets of responses except the principal comments on National Standards following the collapsing of categories, with the first set of principal responses (regarding the impact of National Standards on NZC implementation) now showing an alpha statistic of 0.7.

Some of the variability between coders can potentially be explained by differences in background knowledge and experience relating to National Standards. Apparently differing interpretations need to be considered as, while a coder's knowledge and expertise in their area of research may better equip them to interpret data (in this case, particularly the responses from sector professionals which may at times assume a certain level of prior knowledge), there is also a risk that existing understandings can lead to assumptions that bias results.

There are multiple considerations in deciding on the minimum acceptable level for a reliability statistic. These considerations include the type of statistic used, with higher levels usually needed for more liberal measures, such as proportion of overall agreement, than for more conservative measures, such as Scott's Pi or Krippendorff's Alpha (Lombard et al., 2004). The nature of the variable being coded as well as the number of raters, response categories, and observations being rated or coded are also important in defining acceptable levels (De Swert, 2012). As a general guide, Lombard et al. (2004) suggest that a reliability coefficient of 0.9 or greater is nearly always acceptable, 0.8 and greater is acceptable in most situations, and 0.7 may be an appropriate level of acceptability under some circumstances.

The second and third sets of reliability statistics, which were the ones that represented the data as it was used (with responses assigned to "none" being excluded from analysis), showed a proportion of overall agreement greater than 0.8 for all questionnaire types. The alpha statistics were greater than 0.7 for all but the second set of principal responses. Given that the proportion of overall agreement for the principal responses is above or very close to 0.9, and the alpha statistics are above or close to 0.7; and, given that the analysis is exploratory in nature, reliability is considered sufficiently high to support use of the coded data.

4.3 Considering whether to use weighting

Wylie and Hodgen (2010) report using weights to ensure representativity of all schools in New Zealand when analysing responses to the NZCER 2010 Primary and Intermediate Schools National Survey. For the purposes of secondary analysis, it was necessary to consider whether to take a similar approach, or to use unweighted responses. The following tables show characteristics of the schools in the population, the schools in the actual sample used in the NZCER 2010 Primary and Intermediate Schools National Survey, and the schools in the sample after weighting. In each case, the "%" column shows the percentage of schools in that category. Percentages are given to 2 decimal places. This information was obtained from records kept by NZCER pertaining to the initial analysis.

Decile band by Size

Population	%	Sample	%	Weighted sample	%
low*large	3.18	low*large	5.71	low*large	3.18
low*medium-large	4.02	low*medium-large	4	low*medium-large	4.02
low*small-medium	5.69	low*small-medium	5.43	low*small-medium	5.69
low*small	5.43	low*small	2	low*small	5.43
mid*large	11.11	mid*large	17.71	mid*large	11.11
mid*medium-large	14.08	mid*medium-large	18.29	mid*medium-large	14.08
mid*small-medium	12.83	mid*small-medium	11.43	mid*small-medium	12.83
mid*small	22.17	mid*small	8	mid*small	22.17
high*large	7.09	high*large	14	high*large	7.09
high*medium-large	4.96	high*medium-large	7.43	high*medium-large	4.96
high*small-medium	4.33	high*small-medium	4	high*small-medium	4.33
high*small	5.11	high*small	2	high*small	5.11

Urban and Rural

Population	%	Sample	%	Weighted Sample	%
Rural	33.12	Rural	19.71	Rural	35.17
Urban	66.88	Urban	80.29	Urban	64.83

School Type

Population	%	Sample	%	Weighted Sample	%
Contributing	40.53	Contributing	53.43	Contributing	45.8
Full Primary	53.16	Full Primary	39.14	Full Primary	49.54
Intermediate	6.31	Intermediate	7.43	Intermediate	4.66

Proportion Māori

Population	%	Sample	%	Weighted Sample	%
<8%	15.91	<8%	18.57	<8%	15.61
8-15%	24.41	8-15%	26.29	8-15%	22.95
16-29%	27.07	16-29%	29.14	16-29%	29.5
30-100%	30.93	30-100%	25.71	30-100%	31.15

Proportion Pasifika

Population	%	Sample	%	Weighted Sample	%
<11%	82.89	<11%	79.14	<11%	81.09
>11%	17.11	>11%	20.86	>11%	18.91

Chi-square tests indicated that, at a 5% significance level, the unweighted sample is different from the population in terms of size by decile, rural and urban, and school type.

In particular, rural schools are under-represented, and contributing and intermediate schools are over-represented. High-decile schools are over-represented, and mid-decile schools are under-represented overall. Small schools from all decile bands are under-represented and large schools from all decile bands are over-represented. Medium-large mid-decile schools are also slightly over-represented.

As weighting was based on size and decile, observed values are the same as expected in the weighted sample in relation to those characteristics. None of the other characteristics of the weighted sample are different from those of the population at a 5% level of significance, although there is some indication of difference for school type which has a p-value of 0.09 (1 s.f.). In the weighted sample, it would seem that intermediate schools (and to a lesser extent, full primary schools) are under-represented, while contributing schools are over-represented.

One-way tables showing observed (actual or weighted sample) versus expected (population) frequencies of school type, and associated p-values based on a chi-square goodness-of-fit test are presented here by way of example.

School Type – actual sample

Type	Observed (weighted sample)	Expected (population)	$\frac{(O-E)^2}{E}$ (2 d.p.)
Contributing	187	142	14.26
Full Primary	137	186	12.91
Intermediate	26	22	0.73
Total	350	350	

$$\chi_2^2 = \sum \frac{(O-E)^2}{E} = 27.90$$

p = <0.001

School Type – weighted sample

Type	Observed (weighted sample)	Expected (population)	$\frac{(O-E)^2}{E}$ (2 d.p.)
Contributing	160	142	2.28
Full Primary	173	186	0.91
Intermediate	16	22	1.64
Total	350	350	

$$\chi_2^2 = \sum \frac{(O-E)^2}{E} = 4.83$$

p = 0.09 (1 s.f.)

The weighting of the sample did make significant differences in relation to the representativity of the responses. The secondary analysis presented in this chapter, however, is exploratory in nature, and the results are not intended to be generalisable, particularly as the open responses essentially represent a self-selected sample of the respondents. Therefore, representativity is not a relevant goal, and weighting was not carried out for the purposes of this secondary analysis.

4.4 Methods for statistical analysis

Three methods of statistical analysis were used to explore the binary coding categories of open responses and their relationship with other elements of the survey data. These methods were logistic regression, exploratory factor analysis, and Fisher's exact tests.

4.4.1 Logistic Regression

Logistic regression is a method for predicting the probability of membership of one of two dependent variable categories based on one or more independent variables. The aim of logistic regression is to predict the likelihood that Y (the dependent variable) will be equal to 1 (rather than to 0) given certain values of X (Burns & Burns, 2009).

Logistic regression allows for a combination of categorical and continuous independent variables. It also allows for testing of the effects of a number of independent variables

simultaneously (Burns & Burns, 2009), providing information as to the relative importance of each independent variable as a predictor, and any interaction between the independent variables (Zeigler-Hill, 2013).

Unlike linear regression, logistic regression does not assume a linear relationship between the dependent and independent variables. In logistic regression, a logarithmic transformation is used in order that a non-linear association can be modelled in a linear way (Zeigler-Hill, 2013). Logistic regression does assume a dichotomous dependent variable, with mutually exclusive and exhaustive categories (Burns & Burns, 2009). Other assumptions are the absence of multicollinearity, the absence of outliers, independence of observations, and sufficient ratio of cases to independent variables (too many 0 cell counts will preclude model fit) (Zeigler-Hill, 2013).

The equation for a logistic regression model takes the form:

$$\text{logit}(p) = a + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k$$

(Burns & Burns, 2009). In this equation, a and b_1 to b_k are coefficients, x_1 to x_k are the predictor variables, logit(p) is the natural log of $\frac{p}{(1-p)}$ (the log of the odds), and p is the probability that the dependent variable equals 1.

The predictor (independent) variables can be entered into the model in a range of ways: simultaneously, where all variables are entered at the same time; hierarchically, where variables are entered in blocks that are based on previous research or the theory being tested; or stepwise (including forward selection and backward elimination) where variables are entered or removed on the basis of statistical criteria (Zeigler-Hill, 2013). When carrying out logistic regression using the software package, SAS, stepwise is used to refer to a selection method distinct from forward selection and backward elimination, but combining elements of the two, in that variables can be both entered into and removed from the model, based on the significance of the coefficients (SAS Institute, 2009).

Coefficients, in logistic regression, are estimated on the basis of the log of the odds ratio. The odds ratio represents the relationship between the probability of Y=1 and different values of X. Where the odds ratio is 1, this indicates no relationship between X and Y – the probability that Y=1 does not vary with X. Where the odds ratio is less than 1, a negative relationship is indicated (as X increases, the probability of Y=1 decreases), and where the odds ratio is

greater than 1, a positive relationship is indicated (as X increases, so does the probability of Y=1) (Newsom, 2012). Odds ratios range from 0 to positive infinity. The probability that Y=1 given X is not a linear function, but a logistic function. Therefore, the log-odds is a linear function of the independent variables (Zeigler-Hill, 2013).

Maximum likelihood estimation is used to maximise the probability of classifying cases appropriately given the regression coefficients. It involves an iterative procedure whereby coefficient estimates are varied in order to find the point at which the observed data is most likely (UCLA Statistical Consulting Group, n.d. (1)). This point is where the deviance between observed and predicted values is as small as possible. The deviance statistic is -2 times log-likelihood (sometimes referred to as D) (Newsom, 2012).

To judge the overall fit of the model, the deviance when only the intercept is included is compared with the deviance when the predictors are added. The difference between those values is the likelihood ratio statistic (often referred to as G for goodness of fit) (Newsom, 2012). A p-value for the likelihood ratio statistic is calculated on a chi-squared distribution with degrees of freedom equal to the difference in the number of parameters between the two models being compared (White & Doherty, 2007). The p-value gives the probability, under the null hypothesis, of getting a sample result as or more extreme than the one observed (Clark & Randal, 2011), where the null hypothesis is that all the coefficients in the regression equation are equal to zero (Burns & Burns, 2009). If the p-value is considered sufficiently small, then the null hypothesis is rejected in favour of the alternative hypothesis, which is that at least one of the coefficients of the independent variables in the model is significantly different from zero (Burns & Burns, 2009). Hypothesis tests support rather than prove a hypothesis. In drawing a conclusion based on a hypothesis test, there are four possible outcomes. The null hypothesis can be correctly accepted, or correctly rejected. If the null hypothesis is rejected when it is actually true, this is called a Type I error. The probability of a Type I error is equal to the level of significance chosen for the test (the level at which the p-value is considered sufficiently small to reject the null hypothesis). Accepting the null hypothesis when the alternative hypothesis is true is called a Type II error. Reducing the chosen p-value (level of significance) at which the null hypothesis is rejected reduces the chance of making a Type I error, but increases the chance of making a Type II error. Conversely, increasing the level of significance reduces the chance of making a Type II error, but increases the chance of making a Type I error. Choosing a level of significance for a hypothesis test, therefore, involves some trade-off between Type I and Type II errors (Clark

& Randal, 2011). Generally, p-values smaller than .05 (a probability of 5% of obtaining the observed results under the null hypothesis) are considered to indicate that effects are statistically significant (Zeigler-Hill, 2013). However, the circumstances and implications of the test – specifically, the relative importance of Type I and Type II errors – can support the use of a larger, or a smaller, level of significance.

As well as overall model fit, the effects of individual independent variables can be tested using the likelihood ratio (Bewick et al., 2005), the odds ratio, or the Wald statistic (Zeigler-Hill, 2013). Wald statistics are compared with a chi-square distribution based on 1 degree of freedom (Bewick et al., 2005), and have an associated p-value that can be used for determining significance.

4.4.2 Exploratory Factor Analysis

Factor analysis is a statistical technique commonly used in psychology and education, particularly in the analysis of data from self-report studies. It is a multivariate procedure, useful for dimension reduction, and examining relationships between variables. It is also useful for addressing multicollinearity among variables, enabling the use of other forms of analysis for which multicollinearity poses a problem. There are two main types of factor analysis – exploratory factor analysis and confirmatory factor analysis. Exploratory factor analysis is about generating theories or models, while confirmatory factor analysis is for testing theories or models based on existing knowledge and assumptions (Williams et al., 2010).

Exploratory factor analysis does not impose a preconceived structure on the data, but seeks to identify latent constructs (not directly measured or observable) underlying a set of variables (Suhr, 2006), and represented by linear relationships among those variables (Tryfos, 1998). These constructs are estimated as factors that influence responses on observed variables (Suhr, 2006). Identified factors provide a means of explaining variation using a reduced set of hypothetical variables (Taylor, 2004).

Exploratory factor analysis involves a number of assumptions including an interval or ratio scale of measurement, a random sample of objects or individuals being measured, a linear relationship between observed variables, a normal distribution within each observed variable, a bivariate normal distribution within each pair of observed variables, and multivariate normality (Suhr, 2006). While there are a number of recommendations throughout the

literature as to minimum sample sizes and minimum sample to variable ratios necessary for factor analysis to be a statistically valid procedure, MacCallum et al. (1999) suggest that decisions about the adequacy of the sample to produce a valid and reliable outcome are complex and dependent on other aspects of the data such as the level of communality between variables.

The factor analysis model takes the form:

$$X_1 = a_{11}F_1 + a_{12}F_2 + a_{13}F_3 + \ldots + a_{1j}F_j + e_1$$

$$X_2 = a_{21}F_1 + a_{22}F_2 + a_{23}F_3 + \ldots + a_{2j}F_j + e_2$$

$$\ldots$$

$$X_k = a_{k1}F_1 + a_{k2}F_2 + a_{k3}F_3 + \ldots + a_{kj}F_j + e_k$$

where X_1 to X_k are observed variables, F_1 to F_j are factors, and a_{11} to a_{kj} are coefficients. The coefficients are referred to as factor loadings and, when the factors are uncorrelated with each other, show the correlation between each observed variable and factor in the model. The sum of the squares of the factor loadings for each variable gives the proportion of variation in the variable accounted for by the factors (the communality). The error terms, e_1 to e_k , represent the variation in observed variables not explained by the factors. The larger the communality (and the smaller the error) for each variable, the more successful the factor analysis solution. The sum of the squares of the factor loadings for each factor gives the proportion of the variance of all the variables accounted for by that factor (Taylor, 2004).

There are numerous methods for extracting factors, with the two most common being principal components analysis and principal axis factoring. The practical differences between these two methods are often insignificant, particularly when variables have high reliability, or there are a large number of variables (Williams et al., 2010). The mathematical processes used for extracting factors aim to account for as much of the (remaining) variance in the observed variables as possible in each successive factor. A factor matrix is obtained directly from the matrix of correlations between variables (Suhr, 2006). The rows in a factor matrix represent the coordinates for plotting that variable on the factor axes, where the number of dimensions in which a variable is plotted corresponds with the number of factors in the model.

The decision about how many factors to retain in a factor analysis model involves balancing the retention of information about the observed variables with the creation of an economical set of factors, each having a high level of explanatory power (Taylor, 2004). A number of different criteria for making this decision are available. Williams et al. (2010) suggest that multiple criteria be utilised in deciding on the number of factors to be retained. Kaiser's criterion indicates the retention of all factors with an eigenvalue greater than 1 (Suhr, 2006). The eigenvalue of a factor is equal to the sum of its squared factor loadings for all observed variables (UNESCO, n.d.). The scree test is a criterion that gets its name from the similarities between the graphical representation of a scree plot (a plot of eigenvalues) and the profile of a hillside with a collection of rock debris at the bottom. The bend, or "elbow" (Suhr, 2006, p.3), in this plot is the point at which factors below are considered analogous to scree, while factors above are considered to explain substantially more variation and thus are retained (Williams et al., 2010; Suhr, 2006). Interpreting a scree plot can be a subjective process, particularly when the "elbow" in the plot is gentle rather than abrupt (DeVellis, 2012). The cumulative percentage of variance criterion involves the establishment of a threshold in the cumulative percentage of variance (explained) at which point no further factors are extracted. There are suggestions within the literature about where this threshold should be set, but no consensus or consistency in practice (Williams et al., 2010). Similarly, a threshold can be applied to individual factors (rather than to the model as a whole) where factors are retained only if they explain (at least) a predetermined minimum proportion of the total variance (Suhr, 2006).

In addition to the criteria described above, it is also appropriate to consider the interpretability of the factors – whether the variables that load on the same factor appear to share some kind of meaning, and whether variables that load on different factors appear to have some point of difference, for instance (Suhr, 2006). The initial factor pattern is often not easy to interpret. Interpretation is made easier by ensuring that each observed variable has substantial loading on as few factors as possible. This can be accomplished by rotating the factor axes. The simplest rotation is an orthogonal one (Taylor, 2004). In an orthogonal rotation, the axes remain at 90 degrees to each other (Suhr, 2006). Orthogonal rotation produces a factor structure in which factors are uncorrelated (Williams et al., 2010). Rotating the axes through different angles, so they are not at 90 degrees to each other, is an oblique rotation (Suhr, 2006). An oblique rotation produces a factor structure in which factors are correlated (Williams et al., 2010).

Interpretation and labelling of factors involves examination of the variables that load on that factor, and establishing a theme or name for the latent variable (being the factor) (Williams et al., 2010). The variables with the largest loadings are the ones most strongly correlated with the latent variable, and thus can provide insight into the nature of the factor. When factors explain relatively little variance and include a number of apparently dissimilar observed variables with similar loadings, they should be interpreted carefully, and perhaps not taken too seriously as indicators of a latent variable (DeVellis, 2012). The labelling of factors is a subjective process, dependent on interpretation and subsequent definition by the researcher (Williams et al., 2010).

4.4.3 Fisher's Exact Test

When data is classified by two categorical variables, contingency tables are used to test whether the variables are independent, or whether there is some relationship between them (Clark & Randal, 2011). The general form of a contingency table is:

	Variable B				
Variable A	B_1	B_2	…	B_c	Total
A_1	o_{11}	o_{12}	...	o_{1c}	r_1
A_2	o_{21}	o_{22}	...	o_{2c}	r_2
⋮	⋮	⋮	⋱	⋮	⋮
A_r	o_{r1}	o_{r2}	...	o_{rc}	r_r
Total	c_1	c_2	...	c_c	n

(Clark & Randal, 2011, p.277)

where o_{ij} is the observed frequency in the *ith* row and the *jth* column (the *ith* category of variable A and the *jth* category of variable B), r_i is the total of row i, and c_j is the total of column j. The row and column totals add to n, as do the observed frequencies (Clark and Randal, 2011).

The relationship between the variables in a contingency table can be assessed using Pearson's chi-square test (Howell, 2009). The null hypothesis for a Pearson's chi-square test on a contingency table is that variable A and variable B are independent of one another. Under the

null hypothesis, the expected frequency in row i and column j is denoted by e_{ij} and is equal to:

$$\frac{r_i \times c_j}{n}$$

The test statistic for the chi-squared test is:

$$\chi^2 = \sum_{i,j} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

The larger the sum of the squared differences between the expected and the observed cell counts, the larger the test statistic. The test statistic is compared with a chi-squared distribution with (r-1)×(c-1) degrees of freedom; r being the number of rows in the table, and c being the number of columns, and if it falls outside the rejection region (determined by the level of significance chosen for the test – the p-value), then the null hypothesis is rejected in favour of the alternative hypothesis which is that there is a relationship between variable A and variable B. As a larger test statistic indicates a greater probability that the null hypothesis is incorrect, this is a one-sided, upper-tail test (Clark & Randal, 2011).

A chi-square test for a contingency table is only appropriate if a chi-squared distribution is a good approximation of the distribution of the test statistic. In order for this to be the case, a number of assumptions need to be satisfied. The data needs to come from a random sample, variable categories need to be mutually exclusive, and at least 80% of the expected frequencies calculated need to be greater than 5 (Clark & Randal, 2011).

While Pearson's chi-square test uses estimated probability, Fisher's exact test finds the exact probability for a contingency table. Therefore, Fisher's exact test is more accurate than Pearson's chi-square test, particularly when cell counts are small (IBM, 2013). As with Pearson's chi-square, Fisher's exact test is based on the null hypothesis that variable A and variable B are independent of one another (McDonald, 2009) (with the alternative hypothesis being that there is a relationship between the two variables). To test for independence, the conditional probability of obtaining the observed data, given the particular row and column sums of the contingency table, is calculated (Weisstein, n.d.). If the row and column totals are fixed, then the probability of obtaining the observed data is:

$$\frac{(r_1!\,r_2!\ldots r_i!)(c_1!\,c_2!\ldots c_j!)}{n!\,\Pi_{ij}\,o_{ij}!}$$

Contingency tables are then created for all other possible outcomes that would result in the same row and column totals, and conditional probability calculated for each one. The sum of the probabilities for all possible tables is 1 (Weisstein, n.d.).

Fisher's exact test is most commonly used for 2×2 tables. In this case, a p-value is obtained by adding together all the probabilities less than or equal to the probability of the observed table (Telke, 2009). Because Fisher's exact test is based on an exact distribution, and extremes exist at either end of that distribution, the test is two-sided. If the p-value is smaller than the level of significance chosen for the test, then the null hypothesis is rejected in favour of the alternative hypothesis. For tables larger than 2×2, difference in proportion cannot be used as a measure of dependence by which to order the possible outcomes, and so an alternative measure of association such as the likelihood ratio test is employed (Weisstein, n.d.).

4.5 Analysis and results

Access to the NZCER 2010 Primary and Intermediate Schools National Survey dataset was provided by NZCER. Analysis was carried out in SAS with code written by Edith Hodgen of NZCER for the original analysis of the survey data being used, and adapted where necessary, in preparing the data for secondary analysis. PROC LOGISTIC, PROC FREQ, and PROC FACTOR were used for the secondary analysis.

For the purposes of logistic regression, Y=1 when the open response coding category is “some positive”, and Y=0 when the open response coding category is “negative”. Unless otherwise specified, independent variables were entered into the model simultaneously, as well as stepwise, forward, and backward selection being used.

The level of significance chosen for tests involving p-values was 10%. This means that p-values <0.10 were considered significant. This departure from the more commonly used 5% significance level is due to the exploratory nature of the work. The open responses do not represent a random sample and there is little intention to generalise on the basis of results. In this case, it was considered reasonable to increase the chance of a Type I error, in order to reduce the chance of a Type II error, and thereby increase the power of the tests to detect associations. The exception is for forward and stepwise effects selection methods in logistic

regression, which utilise the SAS default of 0.05 as the significance level for entry and removal of effects. Reported p-values have been rounded to 3 decimal places.

There is also an acknowledgement of an inflated probability of Type I error due to repeated testing. Again, the risk was considered acceptable in light of the exploratory approach.

4.5.1 School characteristics

Logistic regression and Fisher's exact tests were both used to investigate the possibility of relationships between the binary response variable of “negative” or “some positive” comments made in relation to National Standards (herein referred to as the attitude variable, and categorised as negative or positive for ease of reference), and school characteristics. The following school characteristics were considered:

- school type (Contributing, Full primary, Intermediate)
- authority (State, State-integrated)
- urban (Urban, Rural)
- decile band (Low, Mid, High)
- size (Small, Small-medium, Medium-large, Large)
- proportion Māori (<8%, 8-15%, 15-30%, >30%)
- proportion Pasifika (<10%, >10%)
- proportion Asian (<10%, >10%)

Neither logistic regression nor Fisher's exact tests yielded any significant effects relating to school characteristics in the case of the teacher or principal responses.

When logistic regression was carried out with the trustee responses, a convergence failure due to quasi-complete separation of data points was indicated. Quasi-complete separation occurs when the response variable separates a predictor variable, or a combination of predictor variables, to a certain degree (UCLA Statistical Consulting Group, n.d. (2)). The result is that the maximum likelihood estimates do not exist, and so the model fit is not valid. Though SAS does not explicitly state which variable(s) are compromising model validity, this information may be gleaned from an examination of the output. Variables with quasi-complete separation usually have large parameter estimates, very large standard errors, and Wald chi-square statistics close to 0, due to the lack of maximum likelihood estimates. The creation of contingency tables showing each predictor variable cross-tabulated with the

response variable may also be helpful, as zero cell counts can imply quasi-complete separation (Allison, 2008). In this case, school type was the problematic predictor variable, with all of the trustees from intermediate schools falling into the negative category of the attitude variable. When type was removed from the model, the convergence criterion was satisfied. Simultaneous, stepwise, forward, and backward selection methods were then used with the new model. With simultaneous selection, proportion Asian was the only significant effect, with a p-value of 0.015 on the Wald chi-square test. The p-value for the overall model fit using stepwise, forward, and backward selection was 0.002. All three effect selection methods resulted in a model containing proportion Māori and proportion Asian. For stepwise selection, p=0.033 for proportion Māori and p=0.011 for proportion Asian. Forward and backward selection gave a p-value of 0.006 for proportion Māori and a p-value of 0.021 for proportion Asian. While no significant relationship was found between trustee attitude and proportion Asian using Fisher's exact test, examination of frequencies shows that the proportion of positive attitude responses from trustees in schools with >10% Asian students (40% of 81) was higher than that of trustees in schools with <10% Asian students (26% of 23). For trustees, the relationship between the attitude variable and proportion Māori was significant at a 10% level according to Fisher's exact test, with a p-value of 0.071. A smaller proportion of trustees from schools with >15% Māori, than trustees from schools with <15% Māori, were categorised as positive. Specifically, cross-tabulation shows that 44% of 27 trustees in schools with >8% Māori students had positive responses, 52% of 27 trustees in schools with 8-15% Māori students had positive responses, 21% of 28 trustees in schools with 15-30% Māori students had positive responses, and 27% of 22 trustees in schools with >30% Māori students had positive responses. Significant results were also obtained using Fisher's exact test for type and decile band. The p-value for attitude by type was 0.041, with trustees from contributing schools having the highest proportion of positive attitude responses (44% of 64 respondents being categorised as positive), followed by trustees from full primary schools (30% of 33 respondents being categorised as positive), and then by trustees from intermediate schools (0% of 7 respondents being categorised as positive). The p-value for attitude by decile band was 0.027. Trustees from high-decile schools had the greatest proportion of positive attitude responses (53% of 34), and mid-decile schools the lowest (26% of 61). There were only 9 trustees from low-decile schools with responses on the attitude variable. Of those 9, 44% were categorised as positive.

Another convergence failure occurred when logistic regression was carried out with parent responses. Examination of frequency tables showed that only 3 out of 134 parents were from rural schools, and those 3 were all categorised as negative. The urban variable (including the categories urban and rural) was therefore removed from the model which subsequently converged. No significant associations between school characteristics and attitude were found among the parent responses through logistic regression or Fisher's exact tests.

4.5.2 Respondent characteristics

Possible relationships between respondent characteristics and attitude were also investigated using the same methods.

One of the characteristics considered was ethnicity. The ethnicity item in all the questionnaires allowed for multiple responses. Fisher's exact tests for ethnicity were done on the basis of 2x2 tables which looked at each ethnicity category individually, comparing ethnicity category selected with ethnicity category not selected. For logistic regression, an alternative way of categorising ethnicity was used. This variable is called priority ethnicity and was developed as part of NZCER's original analysis. Priority ethnicity allows for only one ethnicity per respondent. Where multiple ethnicities are selected by one respondent, the one used is determined through a priority structure, with the order of priority being: Māori, Pacific, Asian, Other, NZ European/Pakeha. For teachers, a priority variable for role was also used for logistic regression (once again, this category was created by NZCER for primary analysis). The priority order for role was: assistant principal/deputy principal, middle management (including curriculum/syndicate leader, receive a management unit, and senior teacher), subject specialist, classroom teacher, other.

Teachers

Teacher respondent characteristics considered using logistic regression were:

- year levels taught (0-1, 2-3, 3-4, 5-6, 7-8, No class)
- number of years spent teaching (<3, 3-5, 6-10, 11-15, >15)
- number of years spent teaching in that school (<3, 3-5, 6-10, 11-15, >15)
- (priority) role (Assistant principal/Deputy principal, Middle management, Subject specialist, Classroom teacher, Other)
- permanence (Permanent, Fixed-term, Relieving)

- hours of employment (Full-time, Part-time)
- hours spent on work outside those timetabled (0, 1-6, 7-10, 11-15, 16-20, 21-25, >25)
- gender (Male, Female)
- age (<30, 30-39, 40-49, 50-59, >59)
- (priority) ethnicity (Māori, Pacific, Asian, Other, NZ European/Pakeha)

Logistic regression using all of the above variables resulted in quasi-complete separation. This was addressed by removing priority role from the model. Subsequently, significant p-values were obtained for number of years spent teaching in that school (0.032) and permanence (0.016) using simultaneous effect selection. When stepwise, forward, and backward selection were used, only permanence was significant with the p value for the likelihood ratio statistic (indicating overall model fit) being 0.015, and the p-value for the Wald chi-square statistic for the variable permanence being 0.009 for all three effect selection methods. Looking at permanence cross-tabulated with number of years spent teaching in that school indicated that the two are (logically) highly correlated. Cross-tabulation of permanence by attitude showed that permanently employed teachers had the smallest proportion of positive attitude responses (11% of 373), while relieving teachers had the highest (67% of 3). Of the teachers who indicated that they were employed on a fixed term basis, 20% (of 41) also had positive responses on the attitude variable. A Fisher's exact test of permanence by attitude returned a p-value of 0.010, suggesting a statistically significant relationship between these variables. While there was no significant relationship between attitude and number of years spent teaching in that school detected using Fisher's exact test, cross-tabulation shows that the proportion of positive attitude responses decreased as years spent teaching in that school increased, with positive attitude responses from: 16% of (128) teachers who had been in the school less than 3 years; 14% of (94) teachers who had been in the school 3-5 years; 11% of (93) teachers who had been in the school 6-10 years; 11% of (46) teachers who had been in the school 11-15 years; and, 4% of (56) teachers who had been in the school for more than 15 years.

A number of variables associated with teacher's perceptions of their job were also investigated, in relation to their association with attitude, using Fisher's exact tests. These variables came from questions allowing for multiple responses, and are too numerous to list. Each response category was treated as an independent variable and analysed using a 2x2 table

and a test based on whether the category was selected or not selected by the respondent. The questions, and associated significant variables, were as follows:

- Question 31: What do you feel are your main achievements as a teacher in the last three years?

Significant relationships were found between attitude and three of the response options for this question. A positive attitude response was more likely among those teacher respondents who selected "Began to use NZC" (p=0.003), "More involvement of parents with students learning" (p=0.013), and "Refined/introduced new assessments" (p=0.010), than among those who did not select these options.

- Question 32: What are the main things you would change about your work as a teacher?

Significant relationships were found for four of the options associated with this question. A positive attitude response was less likely among those who selected "Reduce assessment workload" (p=0.004), "Reduce assessment requirements" (p=0.040), and "More appreciation of my work by school management" (p=0.087); and more likely among those who selected "More advice available when assessment results show gap in student learning" (p=0.038).

- Question 33: Which of the following best reflects your career plans for the next five years?

Teachers who selected "Increase level of responsibility" were more likely to have a positive attitude response than those who did not select this option (p=0.003).

Principals

Principal respondent characteristics considered using logistic regression were:

- number of years as a principal (<3, 3-5, 6-10, 11-15, >15)
- number of years as a principal in that school (<3, 3-5, 6-10, 11-15, >15)
- number of schools in which respondent has been a principal (1, 2, 3, 4, 5, >5)
- years spent teaching before becoming a principal (<3, 3-5, 6-10, 11-15, >15)
- years spent in (school) senior management before becoming a principal (<3, 3-5, 6-10, 11-15, >15)
- hours spent working per week (<41, 41-50, 51-55, 56-60, 61-65, 66-70, 71-80, >80)

- optimism about life and job (Very optimistic, Quite optimistic, Occasionally optimistic, Never)
- typical stress level so far this year (Extremely low, Low, About average, High, Extremely high)
- current state of health (I am exceptionally healthy, I am very healthy and rarely get sick, I am generally healthy, My health is really not good)
- level of tiredness over the past week (Wide awake and rearing to go most of the time, Some level of tiredness through the days, Constant feeling of tiredness that is not affecting my performance, Constant feeling of tiredness that is affecting my performance, Absolutely worn out)
- gender (Male, Female)
- age (<40, 40-49, 50-59, >59)
- (priority) ethnicity (Māori, Pacific, Asian, Other, NZ European/Pakeha)

Quasi-complete separation was detected, and priority ethnicity was removed from the model to address this. After removal of priority ethnicity, logistic regression using the simultaneous selection method gave a significant p-value (0.015) for years spent teaching before becoming a principal. Stepwise, forward, and backward selection also indicated that years spent teaching before becoming a principal was the only significant predictor variable, with the p-value for the likelihood ratio statistic (model fit) being 0.024, and the p-value based on the Wald chi-square statistic (relating to the significance of the effect of the predictor variable) being 0.027 for all three of those selection methods.

Fisher's exact tests were carried out for all of the variables above, with the only significant result being for optimism about life and job (p=0.001). This variable was difficult to interpret, as although it was designed to be ordinal, its relationship with the attitude variable did not show a clear pattern. Among those who answered “Very optimistic”, “Occasionally optimistic”, and “Never”, there were similar proportions of respondents with a positive attitude response (between 33% and 44% in each case). There was a much smaller proportion of principals who selected “Quite optimistic” with a positive attitude response (only 9%). The Fisher's exact test for years spent teaching before becoming a principal did not give a significant p-value using a 10% level of significance (p=0.254). Examination of the frequencies, however, showed that the proportion of principals with a positive attitude response decreased as the number of years spent teaching before becoming a principal

increased. Sixty-seven percent of the 3 principals who taught for <3 years, 36% of the 14 principals who taught for 3-5 years, 24% of the 25 principals who taught for 6-10 years, 23% of the 26 principals who taught for 11-15 years, and 17% of the 36 principals who taught for >15 years, had a positive attitude response. This is likely to be because, while the logistic regression considered the variable categories as ordinal, the Fisher's exact test considered them as nominal, so the ordered nature of the relationship was not taken into account with the exact test.

As with the teacher responses, a number of additional variables were examined using Fisher's exact test in order to determine whether they might be associated with the attitude variable. These additional variables were obtained from questions allowing for multiple responses and were analysed using 2x2 tables, on the basis of whether they were or were not selected. Questions and results were:

- Question 54: What was your last position before you first became a principal?

No significant results were found for any of the options associated with this question.

- Question 57: Have you ever taken another position between principalships?

Principals who selected “Yes – advisor” in response to the first question were more likely to have positive attitude responses than those who did not select this option (p=0.023).

- Question 62: What teaching do you do?

Principals who selected “Model lessons for teachers” were more likely to have a positive attitude response those who did not select this option (p=0.074), while those who selected “None” were less likely to have a positive attitude response than those who did not select “None” (p=0.065).

- Question 66: How do you manage your workload?

No significant results were found for any of the options associated with this question.

- Question 67: What are the main things you would change about your work as a principal?

Principals who selected the option “Have more teaching staff I could delegate to” were more likely to have a positive attitude response than principals who did not select this option (p=0.035).

- Question 68: Which of the following best reflects your career plans for the next five years?

Principals who selected “Apply for a study award/sabbatical/fellowship” were more likely to have a positive attitude response (p=0.040), while principals who selected “Change to a different career” were less likely to have a positive attitude response (p=0.035).

- Question 69: What do you think are your main achievements as a principal in the last three years?

Principals that selected the option “Pasifika student performance levels stayed high or improved” were less likely to have a positive attitude response than those who did not select this option (p=0.076).

Trustees

Trustee respondent characteristics considered using logistic regression were:

- years on this board (<1, 1-1.9, 2-2.9, 3-3.9, 4-4.9, 5-5.9, 6-6.9, 7-7.9, 8-8.9, 9-9.9, >9.9)
- average hours per week spent on trustee role (<2, 2-5, 6-10, >10)
- board chair (Yes, No)
- highest educational qualification (Postgraduate Degree/diploma, Bachelor degree, Undergraduate diploma/certificate or equivalent, Technician's certificate or equivalent, Trade certificate or equivalent, Pre-vocational certificate or equivalent, University entrance or equivalent, Sixth form certificate or equivalent, School certificate or equivalent, None, Other)
- in paid employment (Yes, No)
- gender (Male, Female)
- age (<30, 30-39, 40-49, >49)
- (priority) ethnicity (Māori, Pacific, Asian, Other, NZ European/Pakeha)

Quasi-complete separation was detected, and priority ethnicity was removed from the model to address this. Subsequently convergence was achieved, but no significant effects were detected.

Of the variables above, significant results using Fisher's exact tests were obtained only for average hours per week spent on trustee role (p=0.071). The frequency table for this variable

indicates that spending a greater number of hours per week is generally associated with an increased likelihood of having a positive response on the attitude variable. Twenty-five percent of those spending <2 hours per week, 47% of those spending 2-6 hours per week, and 50% of those spending 6-10 hours per week, had a positive attitude response. This pattern did not hold for trustees spending >10 hours per week on their work. However, there were only 2 respondents in this category (both of whom had a negative attitude response).

Other questions from the trustee questionnaire that allowed for multiple responses were investigated using Fisher's exact test with 2x2 tables (option selected versus option not selected by positive attitude response versus negative attitude response). These questions were:

- Question 2: Have you been a school trustee before?

No significant results were found for any of the options associated with this question.

- Question 3: Have you served on the boards of any other organisation?

No significant results were found for any of the options associated with this question.

- Question 4: Why did you decide to go on this board?

Trustees who selected the option “I wanted to improve achievement levels” were more likely to have a positive attitude response than those who did not select this option (p=0.053).

- Question 5: What have you gained from being on this board?

No significant results were found for any of the options associated with this question.

- Question 7: What is your current role on the board?

No significant results were found for any of the options associated with this question.

- Question 9: What do you think are the key elements in your role as a trustee?

Trustees who selected the option “Supporting school staff/principal” were less likely to have a positive attitude response than those who did not select this option (p=0.094).

- Question 14: What are the main things you would change in your role as a trustee?

Significant relationships were found for four of the options associated with this question. A positive attitude response was more likely among those who selected “No changes”

(p=0.046), and those who selected “Have better information from school staff to inform our decisions” (p=0.058) than among those who did not select these options. However, it is worth noting that the number of respondents selecting these options was very small (3, and 5 respectively). Trustees who selected “Get more guidance on how to use achievement data to inform board decision making” were also more likely to have positive attitude responses than those who did not select this option (p=0.047). Those who selected “Receive more funding for the school” were less likely to have a positive attitude response than those who did not select this option (p=0.022).

- Question 33: What do you feel are your board's main achievements over the last year?

Trustees who selected “Improvements in student behaviour” were less likely to have a positive attitude response (p=0.063), as were those who selected “Improvements in Māori students’ achievement” (p=0.048), than those who did not select these options.

Parents

Parent respondent characteristics considered using logistic regression were:

- number of children at the school (1, 2, 3)
- year level of youngest child at the school (0-1, 2-3, 4-6, 7-8)
- number of years respondent has had a child at this school (1, 2, 3, 4, 5, 6, >6)
- highest educational qualification (Postgraduate degree/diploma, Bachelor degree, Undergraduate diploma/certificate or equivalent, Technician's certificate or equivalent, Trade certificate or equivalent, Pre-vocational certificate or equivalent, University entrance or equivalent, Sixth form certificate or equivalent, School certificate or equivalent, None, Other)
- gender (Male, Female)
- (priority) ethnicity (Māori, Pacific, Asian, Other, NZ European/Pakeha)

Quasi-complete separation was detected, and priority ethnicity was removed from the model. Subsequently convergence was achieved. The p-value associated with the Wald chi-square statistic for highest educational qualification was 0.083. This effect was not retained in the model using any of stepwise, forward, or backward selection (due to the required level of significance used being p=0.05 in accordance with SAS default settings for proc logistic). Examination of the frequency table for this variable did not show any clear pattern in any case. The Fisher's exact test for this variable did not return a significant p-value.

The question relating to children's year levels in the parent questionnaire was a multiple response item, asking parents to select all the year levels of all their children currently at the school. For logistic regression, a variable constructed for NZCER's initial analysis of the data, which included only the lowest year level selected was used. This allowed for analysis of year level as a single variable with multiple categories (the variable then being the year level of youngest child). For Fisher's exact tests, year level was considered as 9 independent variables (Years 0 through 8) and the full set of responses was used. A relationship, significant at a 10% level (p=0.077), was found between the attitude variable and having a child at the school in Year 8, with parents who selected Year 8 being less likely to have a positive attitude response than those who did not select Year 8. Significant results were also obtained in relation to ethnicity, with parents who selected Pacific being less likely to have a positive attitude response than those who did not select Pacific (p=0.076) and those who selected Asian being more likely to have a positive attitude response than those who did not select Asian (p=0.051). It should be noted that the number of respondents selecting these options was relatively small – out of 135 parents, 5 selected Pacific and 11 selected Asian.

Further analyses based on other questions in the parent questionnaire are reported later in this chapter.

4.5.3 National Standards factors from teacher, principal, and trustee data

The teacher, principal, and trustee questionnaires contained a number of Likert-type items (a five point scale with strongly agree to strongly disagree) around the implementation and likely impact of National Standards in the school. For the purposes of NZCER's primary analysis of the data, factor analysis was carried out, and these items were reduced to a smaller number of factors. Instead of using the estimated factors in the analysis, means were created from the items associated with each factor, and these means were used as variables. A similar method was utilised for secondary analysis, in order that the variables created through factor analysis could be easily used across datasets. The factors created in secondary analysis were, in some cases, very similar to those created through primary analysis, while in others, a slightly different process (arising from variation in purpose and foci) led to differences in the make-up of factors.

In extracting factors, eigenvalues and scree plots were examined, and orthogonal rotation was used. Issues involving the interpretability and practicality of using the factors were also considered. In some cases, items that did not appear consistent with the underlying theme,

represented by the group of items as a whole, were excluded, and where a factor did not appear to have a coherent theme (or where the theme it contained was not considered relevant), it was not used. A summary of the National Standards factors created follows, together with the results of logistic regression analyses carried out to explore potential relationships between these factors and respondents' attitude to National Standards, as measured by the attitude variable.

Teachers

The questions included in factor analysis were:

- Question 10: What changes are you seeing now in your work as a result of the introduction of National Standards? (20 items)
- Question 13: How easy have you found it to make OTJs? (6 items)
- Question 15: What have been your experiences of moderating OTJs? (9 items)
- Question 16: Please state your level of agreement with the following statements about the introduction of National Standards. (17 items)
- Question 17: What's your view of the likely impact of the National Standards at your school in the short term? (16 items)

Analysed together, these items produced 18 factors with eigenvalues greater than 1. The scree plot was not easy to interpret as it had a gradual curve rather than an identifiable “elbow”. Different factor analysis models were explored specifying varying numbers of factors. Based on consideration of the shape of the scree plot and the nature of the items loading on each factor, it was decided that 9 factors should be extracted. Two of the nine factors were not used, as they were not considered particularly relevant to the investigation at hand, and a number of individual items were excluded from the remaining 7 factors, in order that each factor represented a coherent theme. Given that means were going to be used in place of factor loadings, items with negative factor loadings were reversed. The resulting factors were:

- Clarity and robustness of standards (comprising 13 items from question 16)
- Positive effects of standards on practice (comprising 4 items from question 10 and 7 items from question 17)
- Time spent on literacy and numeracy (comprising 5 items from question 10)
- Strength of moderation practice (comprising 9 items from question 15)
- Time spent on assessment (comprising 3 items from question 10)

- Ease of making OTJs (comprising 6 items from question 13)
- Influence of use and interpretation of standards on their likely impact (comprising 5 items from question 17)

New variables were created by taking the mean of the items associated with each factor. Logistic regression, with attitude as the response variable, was carried out with the newly created variables. When all the factors were included in the logistic regression model, results indicated that "Positive effects of standards on practice", and "Time spent on assessment" were useful predictors of a positive attitude response. Looking at the variable means by attitude suggests that a positive attitude response is associated with more agreement with statements about the positive effects of standards on practice, and less agreement with statements about an increase in time spent on assessment. When those two factors were taken out of the model, "Clarity and robustness of standards" was the only significant predictor (with a positive attitude response being associated with more agreement that standards are clear and robust). When clarity and robustness of standards was also removed, none of the remaining factors were found to be significant as predictors.

Principals

Questions included in factor analysis were:

- Question 21: Please state your level of agreement with the following statements about the introduction of National Standards. (17 items)
- Question 22: Please state your level of agreement with these statements about the likely impact of the National Standards at your school in the short term? (16 items)

Analysed together, these items produced 9 factors with eigenvalues greater than 1. Different factor analysis models were explored specifying varying numbers of factors. Based on consideration of the shape of the scree plot, the nature of the items loading on each factor, and the degree of loading of those items, it was decided that 7 factors should be extracted. Again, some items were excluded, due to a combination of lack of fit with other items, and low factor loadings. The resulting factors were:

- Clarity of material and guidance (comprising 6 items from question 21)
- Robustness of standards (comprising 3 items from question 21)
- Importance of working across schools (comprising 3 items from question 21)

- Positive effects of standards on practice (comprising 5 items from question 22)
- Influence of use and interpretation of standards on their likely impact (comprising 4 items from question 22)
- Negative impacts on students and parents (comprising 2 items from question 22)
- No impact on achievement (comprising 2 items from question 22)

New variables were created by taking the mean of the items associated with each factor. Logistic regression, with attitude as the response variable, was carried out with the newly created variables. When all the factors were included in the logistic regression model, "Clarity of material and guidance" and "Positive effects of standards on practice" were found to be significant predictors for attitude. Examination of means showed that a positive attitude response was associated with more agreement that material and guidance were clear and consistent, and with more agreement that the introduction of National Standards had had some positive impact on practice. When these two factors were removed from the model, "No impact on achievement" and "Robustness of standards" became significant as predictors. When these factors were also removed, "Negative impacts on students and parents" became significant. No further significant effects were found beyond that point. Positive attitude was associated with less agreement that standards would have no impact on achievement, more agreement that standards were robust, and less agreement that standards would have negative impacts on students and parents.

Trustees

The trustee questionnaire contained only one question with Likert-type items focussed on National Standards. This was:

- Question 38: Please rate your agreement with the following statements about National Standards by circling the appropriate rating. (13 items)

These items produced 4 factors with eigenvalues greater than 1. The scree plot also indicated that 4 factors should be extracted. Looking at the nature of the items loading on each factor, it was decided that only 3 out of the 4 factors were relevant. These were:

- View on implementation and likely impact of standards (comprising 6 items)
- Board's understanding of standards (comprising 3 items)

- Changes in planning and process as a result of standards (comprising 2 items)

Items with negative factor loadings were reversed and new variables were created by taking the mean of the items associated with each factor. Logistic regression, with attitude as the response variable, was carried out with the newly created variables. When all the factors were included in the logistic regression model, "View on implementation and likely impact of standards" was found to be a significant predictor for attitude. Examination of means showed that a positive attitude response was associated with more agreement that implementation and likely impacts were positive. No further significant effects were found.

4.5.4 Parent data

The parent questionnaire did not contain any Likert-type items pertaining to National Standards. In fact, there were few questions, other than the invitation to comment, that were explicitly related to the standards. As a result, no exploration of parents' views on National Standards was undertaken during NZCER's analysis of the data, beyond the coding of the open-responses by theme. The creation of the attitude variable from open responses provides an opportunity to explore some of the potential influences on parents' attitudes.

In looking for possible predictors of parents' attitudes to National Standards (represented by the attitude variable), every item in the parent questionnaire was considered. Results of analyses relating to parent respondent characteristics have been reported earlier in this chapter. This section addresses the remainder of the items. Factor analysis was used with Likert-type items, then logistic regression was carried out with the resulting factors. In the case of the parent questionnaire, factors were not created by calculating the mean of items loading on each factor. Instead, the estimated factors themselves were used. Questions that asked respondents to select one category from a list, or to provide a number (for instance, the number of children the respondent has at the school) were analysed using logistic regression and Fisher's exact tests when the predictor variable categories were ordinal, and just with Fisher's exact tests when categories were nominal. Potential relationships between multiple response items and the attitude variable were investigated by using Fisher's exact test with each of the items. For reference, the parent questionnaire is included as Appendix 1. Given the large number of items, only significant findings from the parent questionnaire are reported. Questions and associated items with significant results were as follows:

Question 10ii: If yes or not sure (to question 10i: Is there is anything you would like to change about your youngest child's education at this school?), please indicate the changes you would like

- More assessment (Fisher's exact test – p=0.077)
- More accountability (Fisher's exact test – p=0.035)
- More project work (Fisher's exact test – p=0.034)
- Greater range of extracurricular activities (Fisher's exact test – p=0.054)

In each case, parents who selected the item were more likely to have a positive attitude response than those who did not.

Question 12: How did you get information about your child's mid-year progress this year?

- Written report (Fisher's exact test – p=0.064)

Parents who selected this item were less likely to have a positive attitude response than those who did not.

Question 13: What information did you get about your child's progress this year?

- Clear information about where they are in relation to the new National Standards in mathematics (Fisher's exact test – p=0.053)
- Clear information about their attitudes/behaviour at school (Fisher's exact test – p=0.097)

Parents who selected the first item were more likely to have a positive attitude response than those who did not, while parents who selected the second item were less likely to have a positive attitude response than those who did not.

Question 15: What are your main sources of information about education other than the school?

- Other parents (Fisher's exact test – p=0.008)
- TV (Fisher's exact test – p=0.038)
- Books (Fisher's exact test – p=0.042)
- Team-Up website (Fisher's exact test – p=0.066)
- Education Review Office (ERO) (Fisher's exact test – p=0.020)

In each case, parents who selected the item were more likely to have a positive attitude response than those who did not.

Question 17i: Is there any information you would like to have about this school that you don't already have?

Of the parents who answered “Yes” to this question, 61% had a positive attitude response, and 39% a negative attitude response. Of the parents who answered “No”, 38% had a positive attitude response and 62% a negative attitude response. Of those who answered “Not sure”, 26% had a positive attitude response, and 74% a negative attitude response (Fisher's exact test – p=0.023).

Question 17ii: If yes or not sure (to Question 17i above), what other information would you like?

- School planning (Fisher's exact test – p=0.011)
- Information in general (Fisher's exact test – p=0.091)
- Overall student achievement (Fisher's exact test – p=0.010)

In each case, parents who selected the item were more likely to have a positive attitude response than those who did not.

Question 18ii: If yes or not sure (to Question 18i: Is there any area of school life you would like to have a say and feel you cannot?), in what area would you like to have more say?

- Child's class/teacher (Fisher's exact test – p=0.097)

Parents who selected this item were more likely to have a positive attitude response than those who did not.

Question 21: Are you satisfied with the way the school develops its charter and annual plan?

In response to this question, the majority of parents answered “Yes”, or “Don't know what is happening”. Those who answered “Yes” were more likely to have a negative attitude response (53% negative compared with 47% positive) while those who answered “Don't know what is happening” were more likely to have a positive attitude response (47% negative compared with 53% positive). However, these proportions are most likely too similar to represent a relationship. There were bigger differences in proportions of negative and positive attitude response among those answering “No – would like more input”, “No – would like

less consultation", "Not really interested", and "Not sure", with a greater proportion of parents with negative attitude responses than positive attitude responses in each of these categories, though the total number of respondents in some categories was very small (particularly "No – would like more input" and "No – would like less consultation" with 3 and 1 respondent(s) respectively).

Finally, some potential influences on parents' attitudes to National Standards from outside the parent dataset were explored. These were: the views and attitudes of stakeholders within the school (being teachers, principals, and trustees); and, what work the school had done in relation to reporting on National Standards (given that student reports constitute parents' main source of interaction with the standards).

In order to investigate how the views of stakeholders within the school might impact on the attitudes of parents, the National Standards factors from the teacher, principal, and trustee surveys were imported into the parent data set and logistic regression was carried out, with parent attitude as the response variable, and the National Standards factors from each of the other stakeholder groups, in turn, as the predictor variables. With multiple teacher respondents per school in most cases, the value of each teacher factor was obtained by taking the mean of the factor values of all the teachers from that school. The same thing was done for trustees, where there were up to two values for each factor per school. Each stakeholder group had to be examined separately, as including them all in the model at once led to quasi-complete separation, given that, within each school, the values of (and therefore relationship between) the principal, teacher, and Trustee factors were fixed. This process did not lead to significant findings for any of the factors.

To look for possible relationships between schools' National Standards reporting practices and parents' attitudes, a number of variables from the principal dataset were used. It was decided that principals' reporting-related responses should be used (rather than teachers') as there was only one set of principal responses per school, avoiding any issues of inconsistent accounts between respondents from the same school, and logistical issues of combining multiple responses into one school-based variable. Accordingly, the following questions from the principals' questionnaire were examined, using Fisher's exact test, for any relationship with parent attitude:

- Question 15: Has your school started work on implementing the National Standards yet?

Only 2 parents were in schools where implementation had not yet begun. Both of them had negative comments, but the numbers are too small to draw any conclusions. This was reflected in the p-value which did not indicate a significant relationship.

- Question 16i: Has your school undertaken work yet in relation to making overall teacher judgements (OTJs) against the National Standards?

No significant relationships were found between parent attitude response and the school principal's response to this question.

- Question 16ii: If yes (to Question 16i), please indicate what work you have done?

Only three items from this question were directly relevant to reporting. Those items asked whether OTJs had been used in mid-year reporting to parents – for reading, writing, or mathematics. None of them were significantly associated with parent attitude.

- Question 17: How did you report progress in relation to the National Standards in this year's mid-year report to parents/whānau?

No significant relationships were found between parent attitude response and the school principal's response to this question.

- Question 18: What format did you use for your first written report to parents this year?

One item from this question yielded significant results. Parents in schools with a principal who indicated that a new format had been designed were less likely to have a positive attitude response than parents in schools with a principal who did not select this option ($p=0.059$).

5 DISCUSSION AND CONCLUSIONS

The thematic coding of the open responses from the NZCER 2010 Primary and Intermediate Schools National Survey described in section 4.1 highlighted a range of concerns surrounding the implementation of National Standards. Many of these same concerns were being raised by sector groups and academics as part of the debate over the National Standards policy that was occurring around the time the survey was undertaken, and are reflected in a parliamentary research paper published in June 2010 (Parliamentary Library, 2010). The issues identified in the research paper include: the speed of implementation and lack of trialling; potentially inadequate professional development and resourcing; potential for inconsistency in teacher judgements; possible detrimental effects of labelling on students; possible detrimental effects of league tables on schools; and, the potential for narrowing of the curriculum in order to focus on standards attainment.

A number of the themes identified in section 4.1 are also present in the final report from the *Research, Analysis and Insight into National Standards* (RAINS) project (Thrupp & White, 2013). This research was commissioned by NZEI and carried out between 2010 and 2013. The aim of the project was to investigate the enactment of National Standards in six case study schools, and consider whether New Zealand's approach to performance standards would avoid the negative effects of high-stakes assessment systems that have been causing some concern elsewhere (for instance, England, the United States, and Australia). In particular, the final RAINS report discussed concerns over increased staff workloads, narrowing of the curriculum, and potential impacts of labelling students. Among the key findings from the project is a suggestion that schools with a relatively high proportion of students from low socio-economic backgrounds, or with special educational needs, are more likely to experience detriment resulting from National Standards. The RAINS project also found some evidence of positive impacts from the National Standards, in improving teachers' understanding of the New Zealand Curriculum levels, and increasing targeting of interventions for students needing additional support. Similarly, the 2012 report from the School Sample Monitoring and Evaluation Project (SSMEP), commissioned by the Ministry of Education to evaluate the implementation of National Standards in schools, noted that many teachers had reported becoming more systematic in their collection of assessment evidence, and developing a better understanding of curriculum achievement expectations. However, the SSMEP report also concluded that Overall Teacher Judgements lacked

dependability, evidenced by inaccuracy and inconsistency across time and between schools (Ward & Thomas, 2013).

As well as being largely consistent with the findings of other research into the implementation of National Standards, the thematic analysis described in section 4.1 is generally supported by other findings of the original NZCER analysis of the 2010 survey data (Wylie & Hodgen, 2010), with many of the coded themes being quite closely related to selected response questions in other parts of the surveys, particularly the principal and teacher versions.

While attitude factors were created for principals, teachers, and trustees as part of the original NZCER analysis (using banks of Likert-type items), the creation of an attitude variable based on open responses was something that had not previously been done. The school-based demographic characteristics found to be significantly associated with the attitude variable in the secondary analysis were different from those found to be significantly associated with attitude factors in the original analysis (Wylie & Hodgen, 2010). This is not necessarily surprising given that open responses able to be coded as positive or negative came from only a subset of the respondents. However, attitude factors created from Likert-type responses were found to be effective predictors for the open response-based attitude variable, suggesting that it was an internally consistent measure.

The findings of analysis using the binary-coded open responses as an attitude variable are subject to limitations which have been outlined in chapter 4. These limitations include the self-selected nature of the sample of those providing open responses, the increased likelihood of Type I error due to repeated testing, and the inherently subjective nature of the coding process. These limitations preclude any generalisation of findings to the wider population. However, some of the results provide potentially interesting material for considering what might underlie or influence the attitudes and experiences of stakeholders as they interact with National Standards.

Statistically significant relationships between the attitude variable and other survey items suggested a possible link between attitude to National Standards and the work history and future career intentions of teachers and principals. Teachers who have more job permanence and who have spent longer teaching in the school, and principals who spent a greater number of years teaching prior to becoming a principal, were more negative in their attitudes to the National Standards. This pattern could be related to a resistance to change, with people who

have been more immersed or invested in a certain culture or way of doing things being more reluctant to embrace changes in the status quo. Principals with prior experience in advisory roles may be more familiar with the workings of education policy outside the classroom, and presumably more amenable to the idea of policy driving practice, if they have chosen to engage in this field. That teachers who are seeking change in their level of responsibility, and principals who are looking to make changes to their own practice through further study, would be more open to the idea of change at a system level is not inconsistent with these ideas. For teachers, wanting more advice about what to do when assessment results show a gap in learning also suggests an openness to change in practice, and a perceived need for improvement. Teachers who counted a greater involvement of parents with student learning and introducing or refining assessments among their main achievements in recent years would appear to value the changes that the National Standards have been intended to bring about. The fact that teachers who reported beginning to use the New Zealand Curriculum among their main achievements were also more positive about the standards might also suggest that teachers more open to development and change are less likely to be resistant to the National Standards. Alternatively, rather than being related to openness or resistance to change, these relationships might be explained in terms of experience. Less experienced teachers (and principals) may be generally less well equipped, and therefore more likely to seek out potential means of improving their own practice, regardless of whether those means are seen to constitute system-level improvement. Similarly, more experienced educators may be better equipped to recognise potential problems that might arise from the use of standards.

Trustees' attitudes to National Standards appeared to be associated with their perception of the role of a trustee. Trustees who indicated that one of their primary motivations for joining the board was to improve achievement levels in the school, and trustees who wanted more guidance on how to use achievement data to inform decision making, were more positive about the National Standards. Trustees who felt one of the key elements in their role was to support the school staff and principal were more negative about National Standards. One possible interpretation is that those who see a greater need for monitoring and management of school performance, and position a trustees role in relation to this need, might be more likely to see value in the National Standards, while those who place greater trust in the professional judgement of educators, and see the role of a trustee as being primarily supportive might be more likely to see this kind of monitoring as unnecessary. Another possibility is that a trustee's view on the importance of school-based achievement levels affects their attitude

toward the National Standards. If a trustee sees attainment of achievement levels as a priority, relative to other aspects of education, they may be more likely to see potential for the National Standards to assist them in fulfilling their role. A trustee who sees a school's function in more holistic terms might not perceive the same value in the standards, especially if they are seen to detract from other valuable education goals through placing an undue emphasis on attainment of achievement benchmarks.

Unlike the teacher, principal, and trustee surveys; there were no banks of Likert-type items in the parent survey component of the NZCER 2010 National Survey from which to gauge parents' views, experiences, and understanding of the standards (see Appendix 1). In general, there appears to be little New Zealand research looking into the views and attitudes of parents toward National Standards. While both the NZCER 2010 National Survey, and the SSMEP Survey of Parents and Whānau asked some questions about the nature and clarity of schools' reporting, wider views and attitudes are ascertained only through examination of comments (Wylie & Hodgen, 2010; Thomas & Ward, 2011). The binary coding of the open responses from NZCER 2010 National Survey, and the subsequent creation of an attitude variable, afforded an opportunity to further explore these views in a different way. Given the relative paucity of information about how parents see and understand the National Standards, a more comprehensive secondary analysis was carried out with the parent questionnaire, than with the other stakeholder questionnaires. The remainder of this discussion is focussed on exploring parent responses, with some of the salient points from both the thematic analysis, and the use of the binary attitude variable, outlined and discussed in the context of related research.

The drive for accountability in education has a central facet: the notion that parents want and deserve information that can help them make choices about which school to send their child to, or to allow them to ask questions of their child's school when the school is not fulfilling expectations. It is also posited that providing parents with quality and accessible assessment information will help to engage them in their child's learning and thereby lift achievement (Ministry of Education, 2010 (3)). Though the claims about what parents want from schools' reporting that constituted an important element of the justification for the introduction of National Standards (National Party, 2008, National Party, n.d.) appear to have been based primarily on anecdote, a piece of research into parent information needs was commissioned by the Ministry of Education and carried out by Colmar Brunton in 2012 (Colmar Brunton, 2012). Using online forum and focus group discussions, the stated purpose of the research

was to explore the information needs of parents, family, and whānau in relation to school selection; school-wide information about learning and wellbeing; and, student progress, achievement, next steps, and wellbeing.

The resulting report indicates that parents and whānau value the ability to compare their child's achievement with others of the same age, and to compare their school's achievement with other schools. However, they see information about student progress and achievement at the individual and school level as relatively difficult to access or understand. The report notes a general lack of awareness or understanding of National Standards among parents.

A focus on the use of National Standards for comparison of students can be seen in the comments from the NZCER 2010 National Survey. As mentioned in section 4.1, some of these comments indicated confusion among parents over the difference between criterion-referenced and norm-referenced assessments, and the relationship of the National Standards with "average" student achievement. While information on student achievement against the National Standards at the school, regional, and national levels is available from Fairfax media (http://origin-interactives.stuff.co.nz/schoolreport/2014/), and from the Ministry of Education (http://www.educationcounts.govt.nz/statistics/schooling/ national-standards/National_Standards), the standards themselves are not particularly suitable for comparing students. As a criterion-referenced form of assessment, National Standards require comparison of students with a benchmark (a standard), rather than with the distribution of achievement among their peers. The only information available to parents for making comparisons between their child's achievement, and the achievement of other students the same age, is the proportional representation of students across the four National Standards categories ("well below", "below", "at", and "above"). Despite this, it seems that the National Standards have been presented as a vehicle for making comparisons (National Party, 2008, National Party, n.d.). If parents are expecting National Standards to serve this purpose, and finding them ill-equipped to do so, it might go some way toward explaining dissatisfaction with access to achievement information. The same argument can be applied to information on student progress. While the standards were presented as a means to assess and monitor progress (National Party, n.d.), the four-category approach to reporting means the standards are limited in their ability to convey progress except where it constitutes a movement between categories.

In contrast to the findings of the Colmar Brunton (2012) report, the SSMEP Survey of Parents and Whānau (Thomas & Ward, 2011) found most parents indicated that reading, writing, and mathematics achievement information in student reports was easy to understand. Similarly, most parent respondents to the NZCER 2010 Primary and Intermediate Schools National Survey said they received clear information about their child's progress in mid-year reporting (Wylie & Hodgen, 2010). In both cases though, there is a suggestion of some underlying confusion among parents. While the SSMEP survey found that 93% of parent respondents reported receiving information about achievement relative to National Standards in their child's 2010 end of year report, a direct analysis of reports from the schools participating in the SSMEP indicated that only 79% included National Standards information. It is suggested by the report's authors that this may be due to non-response bias, with parents from schools that reported on National Standards being more likely to respond to the survey (Thomas & Ward, 2011). However, a similar phenomenon is observable from the NZCER 2010 National Survey data, with parents in schools that had not yet begun implementing National Standards being just as likely to report receiving clear information about their child's performance in relation to the standards, as those in schools where implementation had begun (Wylie & Hodgen, 2010). The differences in the accounts from schools about reporting practice (or the observed reporting practice in the case of the SSMEP data) and parents' accounts of what reporting contained may be another indication of lack of knowledge and understanding of National Standards.

Given then, that one of the primary justifications for the standards was the need to provide parents with information and engage them more in their child's learning, it seems worthy of note that there may be a lack of understanding and knowledge among this group. It would suggest that, while parents are important stakeholders in the education sector, a lack of knowledge means that they may not be well placed to evaluate education policy, and may be less able to take advantage of benefits offered by policy change if they are unaware of, or misinformed about, the nature of those benefits.

The Colmar Brunton report on parent information needs posited that comments suggested parents' attitudes to National Standards were influenced by what the school had communicated about the standards (Colmar Brunton, 2012). If parents are generally uninformed and confused about the National Standards, this would seem to follow, but it is not necessarily supported by findings from the NZCER 2010 National Survey, and the SSMEP Survey of Parents and Whānau. In the latter, there was no tendency for parents from

the same school to make similar comments, negating any noticeable school-based effect (Thomas & Ward, 2011). The analysis of open response data from the NZCER 2010 National Survey, using the binary attitude variable, found no relationship between the attitudes of school staff and the attitudes of parents with children at that school (see section 4.5.4).

Of the items from the NZCER 2010 National Survey that were found to be significantly associated with the attitude variable, a number appeared fairly self-explanatory. For instance, parents who would like to see more assessment and more accountability in their child's education would logically be more positive about the introduction of the National Standards, as would parents who wanted access to more information about school planning and overall student achievement, as these are generally things that National Standards were intended to deliver. Parents who report utilising the Education Review Office as a source of information might also logically be more inclined to see value in the monitoring and accountability facilitated by the introduction of the National Standards, though it may simply be that parents who want more information about education seek that information from a variety of sources. Colmar Brunton's report on parents' information needs suggests that parents sit on a continuum in terms of their involvement in their child's education, with parents at one end of the continuum seeing education as a partnership between school management, teachers, students, and parents; and parents at the other end of the continuum seeing student learning primarily as the responsibility of schools, with teachers being the experts and parents not needing to be actively involved (Colmar Brunton, 2012). Parents who trust the teacher as expert and see less need for active involvement in their child's education might also see less need to seek out “objective” information in the form of ERO evaluation, or evidence of assessment against a set of nationally consistent benchmarks. There is some evidence of a parallel here with trustees – those who see their role as stakeholders in education being a proactive, rather than reactive (or supportive) one appear potentially more likely to embrace the concept of National Standards. For parents, wanting to have a say in who teaches their child is consistent with a desire for more proactive involvement.

While perhaps not self-explanatory, other significant relationships can potentially be explained by considering contextual factors. Parents with students in Year 8 were found to be more negative about National Standards. According to the final RAINS report (Thrupp & White, 2013), as students progress through primary school, the achievement expectations outlined by the standards become further removed from patterns of actual achievement. A mismatch in expectations versus achievement at Year 8 is supported by a paper published by

NZCER that looked at how student achievement in the mathematics Progressive Achievement Test (PAT) compared with National Standards. Results from an equating exercise suggested that only about 35% of Year 8 students were meeting the mathematics standard (NZCER, 2010). The National Standards results published by the Ministry of Education for 2012 and 2013 paint a far less bleak picture, with around two thirds to three quarters of Year 8 students at or above standards in reading, writing and mathematics in 2012 and 2013. However, there is still an apparent downward trend in the proportion of students meeting standards expectations in writing and mathematics from Year 1 to Year 8 (Ministry of Education, n.d. (5)).

The apparent disjuncture between the results of the PAT equating exercise and the mathematics National Standards results from 2012 and 2013 raises some questions, particularly when considered in light of concerns about the subjectivity of OTJs, and the SSMEP finding that OTJs lacked dependability (Ward & Thomas, 2013). While those questions are beyond the scope of this thesis, general perceptions of misalignment between standards and patterns of achievement, or between standards and other commonly used assessment tools are not. These perceptions were present in the thematic analysis of open response data (see section 4.1), and may be exacerbated by, or contributing to, confusion or misunderstanding about different forms of assessment, and what they provide. In any case, where a higher proportion of students are not reaching standards, it is bound to cause more anxiety among stakeholders, which may lead to a more negative attitude overall if the standards are seen as unhelpful or unfair. The relationship between trustees' attitude and school type, which showed trustees from contributing schools were most positive, while trustees from intermediate schools were least positive, might be explained in a similar way.

As noted earlier, schools' reporting is generally parents' main source of interaction with the standards, so some association between parent attitude and questions about reporting is not unexpected. However, interpreting this association is somewhat complex, particularly given some potentially confounding relationships. Parents with children in Year 7 and 8 were more likely to receive a written report about their child's progress against the standards (Wylie & Hodgen, 2010). Secondary analysis found that parents who received a written report were more likely to be negative, according to the constructed attitude variable. This correlation could then be a result of the higher levels of negativity among parents with students in Year 8. Alternatively, parents with students in Year 8 may have been more likely to be negative due to the fact that they were more likely to have received a written report. It is also entirely

possible that there is no causal relationship between attitude and year level, or attitude and reporting formats, and that they are instead, linked to some other underlying variable. With regard to reporting though, there is some related evidence that may support the notion that this had an effect on attitude. For instance, the RAINS research found that the requirement to report at mid-year on progress toward an end of year standard created some tensions. Teachers were reluctant to make a definitive statement about where a student would be at the end of the year, and this reluctance created anxiety for parents (Thrupp & White, 2013). Parent and teacher frustration or dissatisfaction arising from mid-year reporting time lines (both in relation to an end of year standard, and relative to the timing of assessment on which an interim judgement was based which, for formative reasons, was often at the beginning of the year) was also evident in the thematic analysis of comments from the NZCER 2010 National Survey (see section 4.1).

Parents from schools whose principals indicated they had designed a new format for reporting to parents in 2010, were also found to be less likely to be positive about the standards, based on their open responses. Once again, there may be an element of resistance to change, this time on the part of parents. If parents had been used to receiving and interpreting information in a certain way, to receive something unfamiliar may have created uncertainty. Additionally, a new reporting format may require time for fine-tuning by the school, to figure out what works best for all those involved. In any case, if a new reporting format was seen by parents as being less useful to them than what they had previously received, and that new format was associated with the introduction of the National Standards, it could affect their perception of the usefulness of the standards.

Lastly, secondary analysis found that, while parents who reported having received clear information in relation to their child's achievement in relation to the maths standards were more positive, parents who reported receiving clear information about attitude and behaviour were less positive. A logical conclusion to draw in this case, could be that parents who place (relatively) less value on achievement and more value on the role of education in aiding social and emotional development might see (relatively) less value in the standards. Again, there is a possible parallel here with findings from analysis of trustee responses.

As discussed in chapter 2, achievement standards can be located within a broader political discourse. A number of the patterns outlined above appear to support the idea that respondents to the NZCER 2010 National Survey who indicated a positive attitude to the

National Standards through their open responses also seem more likely to be in agreement with other ideas about the purpose of education and the stakeholder roles and relationships that form part of that discourse. This is not unexpected, but highlights the importance of political ideology and its impacts on the experiences of those within the education system. Media coverage of issues in primary education in New Zealand in recent years provides ample evidence of strained, and even adversarial, relationships among stakeholders. A lack of congruence in public expectations, policy intentions, and what schools are able to deliver could only serve to exacerbate this situation, particularly when combined with a highly politicised environment. It would be naive to think that that discord could be eliminated, as while people can generally agree to wanting the best from the education system, if they have fundamentally different ideas about what the education system is for, then they will most certainly come to different conclusions about how it should operate. However, it should be possible to increase peoples' knowledge and understanding of what is being done in primary education, and how it might affect them in practice, thereby reducing their reliance on political rhetoric. It seems obvious that there is a need for more clarity about the National Standards, among parents in particular. The NZCER 2010 National Survey was undertaken during the first year of implementation of the standards, and knowledge and understanding should of course improve with continued exposure. However, other more recent research suggests a persistently high level of confusion, which potentially continues to undermine the working relationships between parents and schools. The NZCER 2013 Primary and Intermediate Schools National Survey asked parents about their understanding of the National Standards, their support for the standards in principle, and their views on whether the standards constitute a valuable record of student learning. Just over half of the parent respondents to the 2013 survey thought they understood the standards, just under half supported them in principle, and just under half thought they provided a valuable record of learning. Comments were made by 29 percent of parents, most of these detailing concerns about various aspects of the standards. The themes identified by NZCER in parent comments from the 2013 survey are very similar to those identified from the 2010 data (Wylie & Bonne, 2014). Two years later, things may have shifted substantially, but it seems unlikely. Should similar questions be asked of parents in the 2016 survey, it will be interesting to see what change has occurred.

6 GLOSSARY

e-asTTle	Online assessment tool to assess student achievement and progress in reading, mathematics and writing
IEA	International Association for the Evaluation of Educational Achievement
NZC	New Zealand Curriculum
NZCER	New Zealand Council for Educational Research
OECD	Organisation for Economic Co-operation and Development
OTJ	Overall Teacher Judgement
PAT	Progressive Achievement Test
PISA	Programme for International Student Assessment
RAINS	Research, Analysis and Insight into National Standards (Project)
SSMEP	School Sample Monitoring and Evaluation Project
TIMSS	Trends in International Mathematics and Science Study
UNESCO	United Nations Educational, Scientific and Cultural Organisation

7 REFERENCES

Adams, R. (2011). *Comments on Kreiner 2011: Is the foundation under PISA solid? A critical look at the scaling model underlying international comparisons of student attainment.* Retrieved from http://www.oecd.org/pisa/47681954.pdf.

Allison, P. D. (2008). Convergence failures in logistic regression. *SAS Global Forum 2008*, paper 360-2008. Retrieved from www2.sas.com/proceedings/forum2008/360-2008.pdf.

Bewick, V., Cheek, L., & Ball, J. (2005). Statistics review 14: Logistic regression. *Critical Care*, 9:1, pp 112–118. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1065119/.

Bieber, T., & Martens, K. (2011). The OECD PISA study as a soft power in education? Lessons from Switzerland and the US. *European Journal of Education*, 46:1, pp 101-116.

Burns, R. P., & Burns, R. (2009). *Business research methods and statistics using SPSS.* Sage Publications Ltd. Retrieved from http://www.uk.sagepub.com/burns/website%20material/Chapter%2024%20-%20Logistic%20regression.pdf.

Chamberlain, M. (2009). The National Standards, *New Zealand Curriculum Updates.* Retrieved from http://nzcurriculum.tki.org.nz/Curriculum-resources/NZC-Updates/NZC-update-2.

Clark, M. J., & Randal, J. A. (2011). *A first course in applied statistics* (2nd ed.). Auckland, New Zealand: Pearson.

Clark, J. (2010). National Standards: The public debate – what was it all about? *New Zealand Journal of Teachers' Work*, 7:2, pp 106-124.

Colmar Brunton (2012). *Research report: Parents, families and whānau and their information needs*. Author.

De Swert, K. (2012). *Calculating inter-coder reliability in media content analysis using Krippendorff's Alpha*. Retrieved from http://www.polcomm.org/wp-content/uploads/ICR01022012.pdf.

DeVellis, R. F. (2012). *Scale development: Theory and applications* (3rd ed.). Thousand Oaks, CA: SAGE Publications.

Figazzolo, L. (2009). *Impact of PISA on the education policy debate*. Retrieved from http://download.ei-ie.org/docs/IRISDocuments/Research%20Website%20Documents/2009-00036-01-E.pdf.

Goldstein, H. (2004). International comparisons of student attainment: some issues arising from the PISA study. *Assessment in Education: Principles, Policy & Practice*, 11:3, pp 319-330.

Guthrie, J. W., & Pierce, L. C. (1990). The international economy and national education reform: a comparison of education reforms in the United States and Great Britain. *Oxford Review of Education*, 16:2, pp 179-205.

Hattie, J. (2011). *National education standards for New Zealand: A research agenda*. Paper presented at the Symposium on Assessment and Learner Outcomes, Wellington, New Zealand. Retrieved from http://www.victoria.ac.nz/education/pdf/JHC-Symposium/John_Hattie.pdf

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1:1, pp 77-89.

Hipkins, R., & Hodgen, E. (2011). *National Standards, moderation challenges and teacher learning*. Paper presented at the Symposium on Assessment and Learner Outcomes. Wellington, September 1-3, 2011. Retrieved from http://www.nzcer.org.nz/system/files/national-standards-moderation-challenges-teacher-learning_0.pdf.

Hopmann, S. T. (2008). No child, no school, no state left behind: Schooling in the age of accountability. *Journal of Curriculum Studies*, 40:4, pp 417-456.

Howell, D. C. (2009). *Chi-square test – analysis of contingency tables*. Retrieved from http://www.uvm.edu/~dhowell/methods8/Supplements/Chi-Square-Folder/ChiSquareTests.pdf.

IBM (2013). *Chi-square test*. Retrieved from http://pic.dhe.ibm.com/infocenter/spssdc/v7r0m0/index.jsp?topic=%2Fcom.spss.ddl%2Fchi_square_test.htm

Kay, M. (2010, February 2). Key kicks off national standards campaign. *The Dominion Post*. Retrieved from http://www.stuff.co.nz/national/education/3283150/Key-kicks-off-education-national-standards-campaign.

Kellaghan, T. (2006). What monitoring mechanisms can be used for cross-national (and national) studies? In K. N. Ross & I. J. Genevois (Eds), *Cross national studies of the quality of education: planning their design and managing their impact* (pp 51-66). Paris, France: International Institute for Educational Planning.

Klenowski, V., & Wyatt-Smith, C. (2010) Standards, teacher judgement and moderation in the context of national curriculum and assessment reform. *Assessment Matters,* 2, pp 107-131.

Kreiner, S. (2011). *Is the foundation under PISA solid? A critical look at the scaling model underlying international comparisons of student attainment.* Research Report 11/1, Dept. of Biostatistics, University of Copenhagen. Retrieved from https://ifsv.sund.ku.dk/biostat/biostat_annualreport/images/c/ca/ResearchReport-2011-1.pdf.

Krippendorff, K. (2004). Reliability in Content Analysis: Some Common Misconceptions and Recommendations. *Human Communication Research*, 30:3, pp 411-433.

Krippendorff, K. (2011). *Computing Krippendorff's Alpha-Reliability*. Retrieved from http://repository.upenn.edu/asc_papers/43.

Lachat, M. A. (1999). *Standards, equity and cultural diversity*. Providence, RI: LAB at Brown University.

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2004). *Practical resources for assessing and reporting intercoder reliability in content analysis research projects*. Retrieved from http://matthewlombard.com/reliability/.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4:1, pp 84-99.

Machin, S. (2008). The new economics of education: Methods, evidence and policy. *Journal of Population Economics*, 21:1, pp 1-19.

McDonald, J. H. (2009). *Handbook of biological statistics* (2nd ed.). Baltimore, Maryland: Sparky House Publishing.

Ministry of Education (2009 (1)). *National Standards information for schools*. Learning Media Ltd. Retrieved from http://nzcurriculum.tki.org.nz/National-Standards/Key-information/Information-for-schools/National-Standards-launch-pack/.

Ministry of Education (2009 (2)). *The role of formal assessment tools*. Retrieved from http://nzcurriculum.tki.org.nz/National-Standards/Key-information/Fact-sheets/Assessment-tools.

Ministry of Education (2009 (3)). *Moderation*. Retrieved from http://nzcurriculum.tki.org.nz/National-Standards/Key-information/Fact-sheets/Moderation.

Ministry of Education (2010 (1)). *National Administration Guidelines (NAGs)*. Retrieved from http://nzcurriculum.tki.org.nz/National-Standards/Key-information/Fact-sheets/NAGs-fact-sheet.

Ministry of Education (2010 (2)). *Designing the Mathematics Standards for Year 1-8*. Retrieved from http://nzcurriculum.tki.org.nz/National-Standards/Key-information/Background-papers.

Ministry of Education (2010 (3)). *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: New Zealand country background report 2010.* Wellington, New Zealand: Ministry of Education.

Ministry of Education (2013). *The National Administration Guidelines (NAGs).* Retrieved from http://www.minedu.govt.nz/theMinistry/EducationInNewZealand/EducationLegislation/TheNationalAdministrationGuidelinesNAGs.aspx#NAG2A.

Ministry of Education (n.d. (1)). *Designing the Reading and Writing Standards for Years 1-8*. Retrieved from http://nzcurriculum.tki.org.nz/National-Standards/Key-information/Background-papers.

Ministry of Education (n.d. (2)). *Responses to National Standards queries from schools*. Retrieved from http://assessment.tki.org.nz/Overall-teacher-judgment/National-Standards-policy-documents.

Ministry of Education (n.d. (3)). *Clarifications about National Standards*. Retrieved from http://assessment.tki.org.nz/Reporting-to-parents-whānau/Clarifications-about-National-Standards.

Ministry of Education (n.d. (4)). *Making an overall teacher judgement*. Retrieved from http://assessment.tki.org.nz/Overall-teacher-judgment/Making-an-overall-teacher-judgment.

Ministry of Education (n.d. (5)). *2013 Achievement Information*. Retrieved from https://www.educationcounts.govt.nz/statistics/schooling/national-standards/National_Standards.

Mons, N. (2009). *Theoretical and real effects of standardised assessment*. Background paper to the study: National testing of pupils in Europe. Retrieved from http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/111EN.pdf.

Multon, K. (2010). Inter-rater reliability. In N. Salkind (Ed.), *Encyclopedia of research design* (pp. 627-629). Thousand Oaks, CA: SAGE Publications.

National Party (2008). *National Standards: Questions & Answers* [Policy brochure]. Wellington: National Party.

National Party (n.d.). *National Standards: Lifting education standards*. Retrieved from http://www.national.org.nz/files/nationalstandards/national_standards.pdf.

Newsom, J. T. (2012). *Logistic regression* [course notes]. Retrieved from http://www.upa.pdx.edu/IOA/newsom/da2/ho_logistic.pdf.

NZCER (2010). *National Standards for mathematics: where to set the bar?* Wellington, New Zealand: New Zealand Council for Educational Research. Retrieved from http://www.nzcer.org.nz/research/publications/national-standards-mathematics-where-set-bar.

NZCER (n.d.). *National Survey*. Retrieved from http://www.nzcer.org.nz/research/national-survey.

OECD (2010). *PISA 2009 results: What makes a school successful? – Resources, policies and practices (Volume IV)*. Retrieved from http://dx.doi.org/10.1787/9789264091559-en.

OECD (n.d.). *PISA FAQ*. Retrieved from http://www.oecd.org/pisa/aboutpisa/pisafaq.htm.

Parliamentary Library (2010). *National Standards*, Parliamentary support: Research Papers. Retrieved from http://www.parliament.nz/en NZ/ParlSupport/ResearchPapers/a/b/e/00PLSocRP10031-National-Standards.htm.

Pigozzi, M. J. (2006). What is the 'quality of education'? (A UNESCO perspective). In K. N. Ross & I. J. Genevois (Eds), *Cross national studies of the quality of education: planning their design and managing their impact* (pp 39-50). Paris, France: International Institute for Educational Planning.

Potter, J., & Levine-Donnerstein, D. (1999). Rethinking validity and reliability in content analysis. *Journal of Applied Communication Research*, 27, pp 258-284.

Ross, K. N., Paviot, L., & Genevois, I. J. (2006). Introduction: The origins and content of the Policy Forum. In K. N. Ross & I. J. Genevois (Eds), *Cross national studies of the quality of education: planning their design and managing their impact* (pp 25-36). Paris, France: International Institute for Educational Planning.

SAS Institute Inc. (2009). *SAS/STAT 9.2 user's guide* (2nd ed.). Cary, NC: SAS Institute Inc.

Sahlberg, P. (2006). Education reform for raising economic competitiveness. *Journal of Educational Change*, 7:4, pp 259-287.

Suhr, D. (2006). Exploratory or confirmatory factor analysis. *SAS Users Group International Conference*, paper 200-31. Retrieved from www2.sas.com/proceedings/sugi31/200-31.pdf.

Tatto, M.T. (2007). Education reform and the global regulation of teachers' education, development and work: A cross-cultural analysis. *International Journal of Educational Research*, 45:4-5, pp 231-241.

Taylor, A. (2004). *A brief introduction to factor analysis*. Retrieved from http://www.researchgate.net/publictopics.PublicPostFileLoader.html?id=53c50b3fd5a3f2140c8b465e&key=8e9cf6e4-178f-4690-af18-6049b45d5104.

Telke, S. (2009). *PubH 6414 Lesson 12: Chi-square test, Fisher's exact test, McNemar's test* [course notes]. Retrieved from http://www.biostat.umn.edu/~susant/Fall09ph6414/Lesson12.pdf.

Thomas, G., & Ward, J. (2011). *National Standards: School Sample Monitoring & Evaluation Project, 2010 – Survey of Parents and Whānau, March 2011.* Wellington, New Zealand: Ministry of Education.

Thrupp, M., Hattie, J., Crooks, T., & Flockton, L. (2009, 23 November). *Open Letter to the Minister of Education* [press release]. Retrieved from http://www.scoop.co.nz/stories/print.html?path=ED0911/S00107.htm.

Thrupp, M., & White, W. (2013). *Research, Analysis and Insight into National Standards (RAINS) Project. Final report: National Standards and the damage done*. Hamilton, New Zealand: Wilf Malcolm Institute of Educational Research.

Tolley, A. (2008, December 12). *Education Bill to raise standards*. Retrieved from http://www.beehive.govt.nz/release/education-bill-raise-standards.

Torrance, H. (2011). Using assessment to drive the reform of schooling: Time to stop pursuing the chimera? *British Journal of Educational Studies*, 59:4, pp 459-485.

Tryfos, P. (1998). *Methods for business analysis and forecasting: Text & cases*. Wiley. Retrieved from http://www.yorku.ca/ptryfos/f1400.pdf.

UCLA Statistical Consulting Group (n.d. (1)). *FAQ: How are the likelihood ratio, Wald, and Lagrange multiplier (score) tests different and/or similar?* Retrieved from http://www.ats.ucla.edu/stat/mult_pkg/faq/general/nested_tests.htm.

UCLA Statistical Consulting Group (n.d. (2)). *FAQ: What is complete or quasi-complete separation in logistic/probit regression and how do we deal with them?* Retrieved from http://www.ats.ucla.edu/stat/mult_pkg/faq/general/complete_separation_logit_models.htm.

Uebersax, J. (2010). *Statistical Methods for Rater and Diagnostic Agreement.* Retrieved from http://www.john-uebersax.com/stat/agree.htm.

UNESCO (2000). *The Dakar Framework for Action*. Paris, France: UNESCO.

UNESCO (n.d.). *Factor analysis.* Retrieved from http://www.unesco.org/webworld/idams/advguide/Chapt6_3.htm.

Ward, J., & Thomas, G. (2013). *National Standards: School Sample Monitoring and Evaluation Project, 2010-2012. Report to the Ministry of Education*. Wellington, New Zealand: Ministry of Education.

Weisstein, E. W. (n.d.). *Fisher's exact test*. Retrieved from http://mathworld.wolfram.com/FishersExactTest.html.

White, G. C., & Doherty, P. F. (2007). *Likelihood ratio tests* [course notes]. Retrieved from http://warnercnr.colostate.edu/~gwhite/fw663/LikelihoodRatioTests.PDF.

Williams, B., Brown, T., & Onsman, A. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine,* 8:3. Retrieved from http://ro.ecu.edu.au/jephc/vol8/iss3/1.

Wylie, C., & Hodgen, E. (2010). *NZCER 2010 Primary and Intermediate Schools National Survey: A snapshot of overall patterns and findings related to National Standards – revised version*. Wellington, New Zealand: New Zealand Council for Educational Research.

Wylie, C., & Bonne, L. (2014). *Primary and intermediate schools in 2013: Main findings from the NZCER national survey.* Wellington, New Zealand: New Zealand Council for Educational Research.

Zeigler-Hill, V. (2013). *Logistic regression* [course notes]. Retrieved from http://www.zeigler-hill.com/uploads/7/7/3/2/7732402/psy_512_logistic_regression.pdf.

APPENDIX 1: 2010 NZCER National Primary and Intermediate Survey - Questionnaire for Parents or Caregivers

NEW ZEALAND COUNCIL FOR EDUCATIONAL RESEARCH
TE RŪNANGA O AOTEAROA MŌ TE RANGAHAU I TE MĀTAURANGA

P O Box 3237, Wellington 6140
New Zealand
Education House
178-182 Willis Street
Telephone: +64 4 384 7939
Fax: +64 4 384 7933

2010 NZCER National Primary and Intermediate Survey

Questionnaire for Parents or Caregivers

Please fill out this questionnaire by ticking the boxes or circling the numbers that apply to you and/or writing in the spaces provided.

1. How many children do you have at this school? _____

2. What year levels are your children in at this school?

☐ [a] New entrants/Year 0 ☐ [b] Year 1 ☐ [c] Year 2 ☐ [d] Year 3 ☐ [e] Year 4
☐ [f] Year 5 ☐ [g] Year 6 ☐ [h] Year 7 ☐ [i] Year 8

3. How many years have you had a child at this school? _____

*If you have more than one child at this school, for the following questions please give information only about your **youngest** child at this school.*

4. Was this school your first choice for your *youngest* child? *[Please tick one only]*

☐ [1] Yes (it's our closest primary school) ☐ [2] Yes (but it's not the closest primary school)
☐ [3] No *[Please go to Q.6]* ☐ [4] Not sure
☐ [5] Other *[Please describe]* _____

5. If *yes*, how did you get into this school? *[Please tick one only]*

☐ [1] We live in the school enrolment zone ☐ [2] Our chosen school had no enrolment zone
☐ [3] We went into the ballot for this school ☐ [4] We met the special character criteria for the school (e.g., it is a Catholic school and our family is Catholic)
☐ [5] Other *[Please describe]* _____

6. If *no*, what prevented your *youngest* child going to the school of your first choice? *[Please tick all that apply]*

☐ [a] Cost ☐ [b] Transport
☐ [c] School zone ☐ [d] School reluctance to cater for child's special needs
☐ [e] My child did not want to attend that school ☐ [f] The school was full
☐ [g] Other *[Please describe]* _____

7. What information did you use to choose this school for your youngest child? *[Please tick all that apply]*

- ☐ [a] Older child went here
- ☐ [b] Child's friends were going here
- ☐ [c] Other children we knew went here
- ☐ [d] Saw it was the closest school
- ☐ [e] Looked at most recent ERO review of school
- ☐ [f] Family members attended
- ☐ [g] Opinions of parents we know
- ☐ [h] Visited school/attended school open day
- ☐ [i] Newspaper story about this school
- ☐ [j] Newspaper story about other schools
- ☐ [k] School website
- ☐ [l] School annual report
- ☐ [m] Early childhood education centre teachers' views
- ☐ [n] Specialist recommendation
- ☐ [o] Other *[Please describe]* _____

8. Please *circle a number* for the option that best shows how you feel about your youngest child's experience at this school.

	Strongly agree	Agree	Neutral/ not sure	Disagree	Strongly disagree
a) My child's teacher motivates him or her to want to learn	1	2	3	4	5
b) The classroom programme meets my child's academic needs	1	2	3	4	5
c) My child's teacher is committed and enthusiastic	1	2	3	4	5
d) My child's teacher is aware of my child's strengths and weaknesses	1	2	3	4	5
e) I am pleased with the progress my child has made this year	1	2	3	4	5
f) My child's teacher provides clear feedback to my child about his or her work	1	2	3	4	5
g) My child finds school work interesting	1	2	3	4	5
h) The classroom programme involves the right amount of challenge for my child	1	2	3	4	5
i) My child's ideas are listened to	1	2	3	4	5
j) My child is encouraged to be a role model and support other students	1	2	3	4	5
k) My child's teacher responds to any concerns that I have	1	2	3	4	5
l) My child's teacher has high expectations for him or her	1	2	3	4	5
m) My child is helped to set realistic learning goals	1	2	3	4	5
n) The cultural identity of my child is recognised and respected	1	2	3	4	5
o) My child is gaining good social skills	1	2	3	4	5
p) My child is gaining good attitudes to work	1	2	3	4	5
q) I feel welcome when I come to the school	1	2	3	4	5
r) I get good ideas from the school about how to help my child's learning	1	2	3	4	5
s) I would recommend this school to other parents	1	2	3	4	5

9. How well do you think this school helps your child to:

	Very well	Well	Not very well	Not at all well
a) Become self-managing—so they have a "can do" attitude and set high expectations for themselves	1	2	3	4
b) Develop relationship skills—so they can get on well with others	1	2	3	4
c) Develop co-operative skills—so they can work well with others in groups and teams	1	2	3	4
d) Develop problem-solving skills and attitudes	1	2	3	4
e) Develop thinking skills—including being able to ask good questions and be reflective	1	2	3	4
f) Develop research skills—so they can find, judge and use information	1	2	3	4
g) Take part in sport and cultural activities	1	2	3	4
h) Discover a range of interests and passions	1	2	3	4

10. i) Is there anything you would like to change about your youngest child's education at this school?

❒ 1 Yes ❒ 2 No ❒ 3 Not sure

ii) If *yes* or *not sure*, please indicate the changes you would like. *[Please tick all that apply]*

❒ a More interesting work
❒ b More challenging work/higher expectations
❒ c More assessment
❒ d Less assessment
❒ e More emphasis on academic work
❒ f Less emphasis on academic work
❒ g More individual help for students
❒ h Smaller class sizes
❒ i More communication about progress
❒ j More use of computers
❒ k More emphasis on values, relationships and social skills
❒ l More teaching resources
❒ m More accountability
❒ n More parent involvement
❒ o More opportunities for students to feed into decisions or make choices
❒ p More opportunities for children to learn about being a citizen through working in the community
❒ q More opportunity to learn about big issues affecting our world like the environment and sustainability
❒ r More focus on children's cultures and experiences in the school learning programme
❒ s Less strict discipline
❒ t Stricter discipline
❒ u More emphasis on students supporting each other
❒ v More project work
❒ w More information to support my child's learning at home
❒ x Greater range of extracurricular activities (e.g., clubs)
❒ y Less homework
❒ z More homework
❒ aa Other *[Please describe]* _____

11. In general, how would you describe the information you get from the school about:

i) Your youngest child's overall learning programme

❒ 1 Very good ❒ 2 Good ❒ 3 Satisfactory ❒ 4 Poor ❒ 5 Very poor

ii) Your youngest child's learning progress

❒ 1 Very good ❒ 2 Good ❒ 3 Satisfactory ❒ 4 Poor ❒ 5 Very poor

12. How did you get information about your child's mid-year progress this year? [*Please tick all that apply*]

- ❒ [a] Written report
- ❒ [b] Discussion with child's teacher at a set time
- ❒ [c] Discussion with child and child's teacher at a set time
- ❒ [d] Other *[Please describe]* _____

13. What information did you get about your child's mid-year progress this year? [*Please tick all that apply*]

- ❒ [a] Clear information about their progress this year
- ❒ [b] Clear information about my child's learning goals for the rest of the year
- ❒ [c] Clear information about where they are in relation to the new National Standards in *reading*
- ❒ [d] Clear information about where they are in relation to the new National Standards in *writing*
- ❒ [e] Clear information about where they are in relation to the new National Standards in *mathematics*
- ❒ [f] Clear information about their progress in science
- ❒ [g] Clear information about their attitudes/behaviour at school
- ❒ [h] Clear information about what the school is doing to help my child achieve their learning goals
- ❒ [i] Helpful ideas for how I can support my child's learning
- ❒ [j] Less information than I got from last year's mid-year report about their learning as a whole
- ❒ [k] Other *[Please describe]* _____

14. Do you have any comments about the new National Standards?

15. What are your main sources of information about education other than the school? *[Please tick all that apply]*

- ❒ [a] Family
- ❒ [b] Friends
- ❒ [c] Other parents
- ❒ [d] TV
- ❒ [e] Radio
- ❒ [f] Newspaper
- ❒ [g] Books
- ❒ [h] Magazines
- ❒ [i] Internet searches
- ❒ [j] Team-Up website
- ❒ [k] Ministry of Education
- ❒ [l] Education Review Office (ERO)
- ❒ [m] Other *[Please describe]* _____

16. Where do you get information about the school? *[Please tick all that apply]*

- ❒ [a] Weekly school newsletters
- ❒ [b] Newsletters every 2–4 weeks
- ❒ [c] Newsletters once a term
- ❒ [d] Annual school report
- ❒ [e] Look at school website at least once a month
- ❒ [f] Look at school website occasionally
- ❒ [g] Latest ERO report
- ❒ [h] Local community newspaper
- ❒ [i] Other parents
- ❒ [j] Other *[Please describe]* _____

12. How did you get information about your child's mid-year progress this year? [*Please tick all that apply*]

- ❒ [a] Written report
- ❒ [b] Discussion with child's teacher at a set time
- ❒ [c] Discussion with child and child's teacher at a set time
- ❒ [d] Other *[Please describe]* _____

13. What information did you get about your child's mid-year progress this year? [*Please tick all that apply*]

- ❒ [a] Clear information about their progress this year
- ❒ [b] Clear information about my child's learning goals for the rest of the year
- ❒ [c] Clear information about where they are in relation to the new National Standards in *reading*
- ❒ [d] Clear information about where they are in relation to the new National Standards in *writing*
- ❒ [e] Clear information about where they are in relation to the new National Standards in *mathematics*
- ❒ [f] Clear information about their progress in science
- ❒ [g] Clear information about their attitudes/behaviour at school
- ❒ [h] Clear information about what the school is doing to help my child achieve their learning goals
- ❒ [i] Helpful ideas for how I can support my child's learning
- ❒ [j] Less information than I got from last year's mid-year report about their learning as a whole
- ❒ [k] Other *[Please describe]* _____

14. Do you have any comments about the new National Standards?

15. What are your main sources of information about education other than the school? *[Please tick all that apply]*

- ❒ [a] Family
- ❒ [b] Friends
- ❒ [c] Other parents
- ❒ [d] TV
- ❒ [e] Radio
- ❒ [f] Newspaper
- ❒ [g] Books
- ❒ [h] Magazines
- ❒ [i] Internet searches
- ❒ [j] Team-Up website
- ❒ [k] Ministry of Education
- ❒ [l] Education Review Office (ERO)
- ❒ [m] Other *[Please describe]* _____

16. Where do you get information about the school? *[Please tick all that apply]*

- ❒ [a] Weekly school newsletters
- ❒ [b] Newsletters every 2–4 weeks
- ❒ [c] Newsletters once a term
- ❒ [d] Annual school report
- ❒ [e] Look at school website at least once a month
- ❒ [f] Look at school website occasionally
- ❒ [g] Latest ERO report
- ❒ [h] Local community newspaper
- ❒ [i] Other parents
- ❒ [j] Other *[Please describe]* _____

20. Please indicate your views on parents' roles in relation to the school.

	Strongly agree	Agree	Neutral/ not sure	Disagree	Strongly disagree
a) Parents/whānau should have the opportunity to be involved in decisions about their child's learning	1	2	3	4	5
b) Parents/whānau should have the opportunity to be involved in decisions about learning in general at their child's school	1	2	3	4	5
c) Parents/whānau should have regular opportunities to discuss in detail their child's progress	1	2	3	4	5
d) The purpose of interactions between school and parents/whānau should be so *parents/whānau* can better support students' learning	1	2	3	4	5
e) The purpose of interactions between school and parents/whānau should be so *the school* can better support students' learning	1	2	3	4	5
f) The school should provide opportunities for parents/ whānau to work with staff or students in learning projects	1	2	3	4	5

21. Are you satisfied with the way the school develops its charter and annual plan? *[Please tick one only]*

❒ 1 Yes ❒ 2 No—would like more input ❒ 3 No—would like less consultation
❒ 4 Don't know what is happening ❒ 5 Not really interested ❒ 6 Not sure
❒ 7 Other *[Please describe]* _____

22. Do you think the school genuinely consults you about new directions/issues?

❒ 1 Yes ❒ 2 No ❒ 3 Not sure

23. Please indicate your understanding of your school community's attitudes around schooling.

	Strongly agree	Agree	Neutral/ not sure	Disagree	Strongly disagree
a) The school and its community value similar things	1	2	3	4	5
b) The community is open to new learning	1	2	3	4	5
c) The community has realistic expectations of what the school can provide	1	2	3	4	5
d) The community trusts the school	1	2	3	4	5
e) The community feels at home in the school	1	2	3	4	5
f) The community is divided/contains groups with conflicting wishes	1	2	3	4	5
g) The community puts a high value on educational success	1	2	3	4	5

24. i) Did you vote in this year's elections for the Board of Trustees? *[Please tick one only]*

❒ 1 Yes *[Go to ii]*
❒ 2 No *[Go to iii]*
❒ 3 There was no election for the board at this school *[Please go to Q.25]*
❒ 4 Came to this school after the elections *[Please go to Q.25]*

ii) If you *did vote*, what influenced your choice? *[Please tick all that apply]*

- ☐ [a] Candidate had shown previous commitment to school
- ☐ [b] Candidate had been on another board
- ☐ [c] Candidate was known to me
- ☐ [d] Candidate seemed to have skills the school needs
- ☐ [e] Candidate had experience in education
- ☐ [f] Nothing—seemed like a lottery
- ☐ [g] Other _____

iii) If you *didn't vote*, why was this? *[Please tick all that apply]*

- ☐ [a] Didn't get the papers in time
- ☐ [b] Wasn't enough information on the candidates to make a decision
- ☐ [c] Didn't seem important who went on the board—all candidates seemed good
- ☐ [d] Didn't get round to it
- ☐ [e] Other _____

25. Do you feel you have enough contact with the school's trustees?

☐ [1] Yes ☐ [2] No ☐ [3] Not sure

26. What do you think are the key element(s) in the role of the Board of Trustees? *[Please tick the most important]*

- ☐ [a] Providing strategic direction for the school
- ☐ [b] Employing principal
- ☐ [c] Supporting school staff/principal
- ☐ [d] Representing parents in the school
- ☐ [e] Agent of government/representing government interests
- ☐ [f] Scrutinising school performance
- ☐ [g] Overseeing principal
- ☐ [h] Overseeing school finances
- ☐ [i] Other *[Please describe]* _____

27. What do you think are the major issues facing your school, if any? *[Please tick the most important]*

- ☐ [a] Funding
- ☐ [b] Student achievement
- ☐ [c] Attracting good teachers
- ☐ [d] Keeping good teachers
- ☐ [e] Student behaviour/discipline
- ☐ [f] Staffing levels
- ☐ [g] Putting the New Zealand Curriculum into effect for this school
- ☐ [h] Getting good-quality professional development for school staff
- ☐ [i] Property maintenance/ development
- ☐ [j] National Standards introduction
- ☐ [k] Parent/community support
- ☐ [l] Motivating students
- ☐ [m] Assessment workload
- ☐ [n] Principal's leadership
- ☐ [o] Quality of teaching
- ☐ [p] Māori student achievement
- ☐ [q] Pasifika student achievement
- ☐ [r] Achieving school targets
- ☐ [s] Assessment driving the curriculum
- ☐ [t] Including students with special needs
- ☐ [u] Getting a good ERO review
- ☐ [v] Responding to cultural diversity
- ☐ [w] Declining school roll
- ☐ [x] School roll is too large
- ☐ [y] BOT quality
- ☐ [z] Adequacy of ICT equipment
- ☐ [aa] Other *[Please describe]* _____

About you

28. What is your *highest* education qualification? *[Please tick one only]*

- ☐ [1] Postgraduate degree or diploma
- ☐ [2] Bachelor degree
- ☐ [3] Undergraduate diploma/certificate, New Zealand diploma, New Zealand certificate, or national diploma
- ☐ [4] Technician's certificate, advanced trade certificate, or national diploma
- ☐ [5] Trade certificates, apprenticeship, national certificate
- ☐ [6] Pre-vocational certificate, bridging certificate, or foundation certificate
- ☐ [7] University Entrance, Higher School Certificate, National Certificate Level 3, or NCEA Level 3
- ☐ [8] Sixth Form Certificate, National Certificate Level 2, or NCEA Level 2
- ☐ [9] School Certificate, National Certificate Level 1, or NCEA Level 1
- ☐ [10] No qualification
- ☐ [11] Other *[Please describe]* _____

29. Are you:

☐ [1] Female *or* ☐ [2] Male

30. Please indicate the ethnic group(s) with which you identify. *[Please tick all that apply]*

☐ [a] Pākehā/European ☐ [b] Māori ☐ [c] Pacific ☐ [d] Asian

☐ [e] Other *[Please describe]* _____

We really appreciate your views!

Please return your survey to NZCER in the freepost envelope enclosed by
<u>Wednesday 18 August</u>