

**Economics and accounting teachers' beliefs about
school-based assessment, the summative and formative
tension: A mixed method study.**

By

Anne Yates

A thesis
submitted to Victoria University of Wellington
in fulfilment of the requirements for the degree of
Doctor of Philosophy

VICTORIA UNIVERSITY OF WELLINGTON

2018

Abstract

The National Certificate of Educational Achievement (NCEA) is New Zealand's main national qualification for senior secondary school students. A key feature of NCEA is that it allows for more than half of students' final NCEA grades to be assessed by teachers during the school year through school-based assessment, known as *internal assessment*. This key role of teachers in awarding qualifications is likely to have an impact on their conceptions of assessment, conceptions of assessment that are not necessarily fixed. In turn, conceptions can influence teaching practice and are likely to have an impact on how teachers implement internal assessment.

This thesis uses an explanatory sequential, mixed methods design to investigate a group of economics and accounting teachers' conceptions of assessment, their practices in relation to NCEA internally-assessed standards, and the influences on those practices. G.T.L. Brown's (2006) Teachers' Conception of Assessment Abridged (TCoAIIIA) Inventory was used to investigate the participants' conceptions of assessment, and interviews were conducted to probe their internal assessment practices and reasons for those practices.

The quantitative data were analysed using confirmatory factor analysis which showed an inadmissible fit to G.T.L. Brown's (2006) model. A subsequent exploratory factor analysis revealed partly similar and partly different dimensions in the data, compared with those previously reported in other studies using the TCoAIIIA. Qualitative data were thematically analysed and when the qualitative and quantitative data were considered together, greater similarities with G.T.L. Brown's model emerged. Participants revealed four overarching conceptions of assessment: *assessment is for learning*; *assessment is for qualifications*; *assessment is for accountability*; and *assessment is detrimental*. This finding has reinforced the view that teachers' conceptions of assessment are ecologically rational in that a distinct conception that probably relates to the role participants play in assessing for NCEA emerged.

Despite this role, participants also adhered to the conception that a primary function of assessment is to improve students' learning; furthermore, they did not support the use of assessment results as a measure of school quality. An implication of this finding is a belief that promoting a school-accountability use of assessment results is likely to be counter-productive to students' learning.

The qualitative findings revealed a complex set of beliefs and practices towards implementing the internally-assessed component of NCEA, and that beliefs were only one influence on teachers' internal assessment practices. Teachers had to balance their beliefs with the systemic realities of NCEA and their school's policy requirements, and articulated a tension between the *improvement* and *accountability* conceptions of assessment. Moderation processes, procedures and policy encouraged teachers into a cycle of safe rather than innovative internal assessment practice, which means that the original vision for NCEA internal assessment is yet to be realised. There are implications of this finding for professional development, leadership of assessment, and initial teacher education.

One such implication is a requirement for professional development that would provide teachers with successful, innovative internal assessment practices, rather than the present approach which focusses on the reliability of marking. Schools' management needs to take a greater role in leading and encouraging pedagogically sound internal assessment, rather than focussing primarily on agreement rates with NZQA moderators. Initial teacher educators could also introduce student teachers to effective internal assessment practices and to encourage such practices. In addition, attempts to change teachers' assessment practices need to consider existing conceptions of assessment because beliefs have an impact on practices and may need to be challenged. While there is debate about whether beliefs change practice or vice versa, one cannot be changed without considering the other.

Acknowledgements

Research cannot take place without willing participants, so my first acknowledgement must go to the 135 teachers of economics and accounting who completed the online surveys and to the 21 who agreed to be interviewed. They willingly gave time from their busy working lives to contribute to this study. It was their journey in implementing the internally-assessed component of NCEA that inspired this research. I would like to acknowledge all teachers in their endeavours to educate our tamariki.

Supervisors are also essential to the successful completion of a doctoral dissertation and their expert guidance and care mean I have finally reached this point. I owe gratitude to Dr Michael Johnston who guided me in the quantitative realm, and for his careful critique of all aspects of this work. I am most grateful to Emerita Professor Luanna Meyer for continuing to support me, and my work, into her retirement from Victoria University, and am very proud to be her final doctoral student.

My colleagues in the Faculty of Education have been a tower of strength, empathising with the enormity of completing a PhD later in one's career. They have offered support through academic discussions and encouragement, and also by their willingness to be an ear for the trials and tribulations that inevitably emerge on the journey. The Faculty of Education has been generous with support of time, research grants, study leave, workshops and access to expertise. I am also grateful for Susan Kaiser's expert help in preparing the final manuscript and thank you to my esteemed examiners, I greatly value your perspectives and critical feedback.

My parents, Alice, and the late George, highly valued education and encouraged me to aim high. My Scottish heritage is reflected in the pseudonyms I chose for the participants, and I believe it is a strong driver in my quest for education. I want to thank my husband, Eric, for his support over the course of this work, for offering helpful editorial suggestions, and for keeping the home fires burning. To my daughter Kenzie, her partner Daniel, my grandsons Jake and Liam, and my sister Judi – thank-you for your love and support throughout this endeavour. I would also like to acknowledge my daughter, the late Heather, whose untimely death created gaps in my motivation, but ultimately spurred me to complete. Jake and Liam – this is for you – aim high and reach the stars in all your endeavours.

Table of Contents

Abstract	i
Acknowledgements	iii
Terms and Abbreviations	x
CHAPTER 1 Introduction and background to the study	1
1.1 The structure of NCEA	2
1.2 Economics and accounting in NCEA.....	6
1.3 Rationale for the study	8
1.4 The researcher's position	10
1.5 Thesis overview	12
CHAPTER 2 Literature review	13
2.1 Defining assessment.....	13
2.2 School-based (internal) assessment.....	15
2.2.1 Issues of validity and reliability.....	19
2.2.2 Validation and moderation	22
2.2.3 Impact of internal assessment on teachers.....	24
2.2.4 Impact of internal assessment on students.....	27
2.2.5 Concluding comment.....	29
2.3 Assessment policy as a tool to improve teaching and learning.....	29
2.3.1 Assessment reform in New Zealand	29
2.3.2 The impact of assessment reform	32
2.4 Teachers' beliefs	35
2.4.1 Measuring teachers' conceptions of assessment.....	37
2.4.2 Differences in teachers' conceptions of assessment.....	39
2.5 Teachers' assessment identity.....	41
2.6 Conclusion.....	43
2.7 Research questions	43
2.8 Significance of the study	44
CHAPTER 3 Methodology	47
3.1. Theoretical positioning of the study.....	47
3.2 Research design.....	47
3.3 Participants.....	49
3.3.1 Participant selection.....	49
3.3.2 Survey participants	49

3.3.3 Interview participants	49
3.3.4 Sample limitations	50
3.4 Data collection methods	53
3.4.1 Self-report questionnaires	53
3.4.2 Survey administration and quantitative data collection	54
3.4.3 Survey qualitative questions	54
3.4.4 Semi-structured interviews	54
3.4.5 Data collection timetable	56
3.5 Ethical considerations	57
3.6 Quantitative data analysis	57
3.6.1. Factor analysis	57
3.7 Qualitative data analysis	63
3.7.1 Data Reduction	64
3.7.2 Data display	64
3.7.3 Drawing and verifying conclusions	65
3.8 Validity and reliability in a mixed methods design	66
3.8.1 Validity and reliability in quantitative data	66
3.8.2 Validity and reliability in qualitative data	67
3.8.3 The inference process	68
CHAPTER 4 Quantitative findings.....	71
4.1 Confirmatory analysis	71
4.1.1 Strength of the present data loadings to factors.....	73
4.2 Acceptability of G.T.L. Brown’s (2006) model.....	76
4.3 Exploratory factor analysis	78
4.3.1 Four factor analysis.....	78
4.3.2 Three factor analysis.....	80
4.3.3 Five factor analysis	81
4.3.4 Removing items	82
4.3.5 Factors extracted	84
4.4 Conclusion	89
CHAPTER 5 Qualitative findings	95
5.1 Gatekeepers and ushers	95
5.2 Thematic analysis of the interviews	100
5.2.1 Assessment as pedagogy.....	101
5.2.2 Assessment for educational outcomes	104
5.2.3 Assessment as practice	111
5.2.4 Assessment as motivator.....	119

5.2.5 Assessment is detrimental.....	123
5.3 Conclusion.....	137
CHAPTER 6 Discussion of the strands.....	139
6.1 Discussion of the quantitative findings	140
6.1.1 Confirmatory factor analysis	140
6.1.2 Exploratory factor analysis	143
6.1.3 Participants' conceptions of assessment.....	148
6.2 Discussion of qualitative themes.....	149
6.2.1 Assessment as pedagogy.....	149
6.2.2 Assessment as motivator.....	150
6.2.3 Assessment for educational outcomes	152
6.2.4 Assessment as practice and is detrimental.....	154
6.2.5 Internal assessment practices	162
CHAPTER 7 Discussion – Converging the data strands.....	165
7.1 Overarching conceptions.....	167
7.1.1 Assessment is for learning	167
7.1.2 Assessment is for qualifications	169
7.1.3 Assessment is detrimental.....	170
7.1.4 Assessment is for accountability	171
7.2 Influences on teachers' internal assessment practice	173
7.2.1 The macro-context	174
7.2.2 The meso-level context.....	175
7.2.3 Beliefs about assessment	177
7.3 Enacting and embedding assessment reform	178
CHAPTER 8 Conclusion	181
8.1 Summary of the study	181
8.2 Research outcomes.....	181
8.2.1 Conceptions of assessment	181
8.2.2 Internal assessment practices	183
8.2.3 The impact of accountability on internal assessment practices	184
8.3 Contributions to the field	184
8.3.1 Conceptions of assessment are ecologically rational.....	184
8.3.2 Contradictory conceptions are rational	185
8.3.3 Vision for NCEA yet to be realised	186
8.3.4 The meso-level is a dominant impact	186
8.3.5 Teachers are responsible in their role as assessors	187
8.3.6 Teachers are vulnerable in their role as assessors.....	187

8.4 Limitations of this study	188
8.4.1 Sampling	188
8.4.2 Instruments	188
8.4.3 Analysis	189
8.5 Implications of this study	190
8.5.1 Changes to internal assessment practice are needed.....	190
8.5.2 Professional development is needed.....	191
8.5.3 Leadership of internal assessment	193
8.5.4 Teacher education.....	194
8.5.5 Assessment policy should work for teachers and students.....	195
8.6 Areas for future research.....	196
8.7 Significance of the thesis	197
References	199
 Appendices	
Appendix 1: The survey	227
Appendix 2: The generic interview schedule.....	236
Appendix 3: Ethics approval.....	237
Appendix 4: Ethics amendment	238
Appendix 5: Factor matrix for 4 factor solution	239
Appendix 6: The anti-image correlation matrix for a four factor solution	240

List of Tables

Table 1.1	Requirements to gain NCEA at each level	3
Table 1.2	Total number of standards and credits available at each level for accounting and economics.	3
Table 1.3	Economics matrix for NCEA (Ministry of Education, n.d.a)	7
Table 1.4	Accounting matrix for NCEA (Ministry of Education, n.d.b)	8
Table 3.1	Demographic characteristics of interviewees	50
Table 3.2	Percentages of participants in demographic categories compared with national figures	51
Table 3.3	Data collection timetable	56
Table 3.4	Range of Acceptable Fit Indices In CFA	60
Table 4.1	Fit Indices for 135 Secondary Teachers of Accounting and Economics	71
Table 4.2	Items relating to first order factors which contribute to second order factor of “Assessment improves education”	73
Table 4.3	Items relating to first order factor of “Assessment Makes Schools Accountable”	74
Table 4.4	Items relating to first order factor of “Assessment Makes Students Accountable”	75
Table 4.5	Items relating to first order factors which contribute to second order factor of “Assessment is Irrelevant”	75
Table 4.6	Conceptions of assessment inter-correlations: New Zealand primary and secondary teachers (G.T.L. Brown, 2011a) compared with economics and accounting teachers	77
Table 4.7	Pattern matrix extracted using Maximum Likelihood Estimation with an Oblique rotation (Promax) with a four-factor solution	79
Table 4.8	Pattern matrix extracted using Maximum Likelihood Estimation with an Oblique rotation (Promax) with a three-factor solution	81
Table 4.9	Pattern matrix extracted using Maximum Likelihood Estimation with an Oblique rotation (Promax) with a four-factor solution and items 2, 11, 18 & 26 removed	83
Table 4.10	Conceptions of assessment inter-correlations: economics accounting teachers	89
Table 4.11	Comparison of items and factors of G.T.L. Brown’s (2006) model with factors and items for economics and accounting teachers	91
Table 5.1	Typology of teachers’ beliefs and actions towards NCEA internal assessment	96
Table 5.2	Themes and their general description	101
Table 7.1	Convergence and complements among quantitative factors and qualitative themes	166

List of Figures

Figure 3.1.	Overview of themes derived from qualitative data.....	64
Figure 3.2.	Example from matrix of themes and some codes contributing to that theme.....	65
Figure 3.3.	Typology of teachers' beliefs and actions towards NCEA internal assessment.....	66
Figure 3.4.	Convergence and complements among quantitative factors and qualitative themes	70
Figure 4.1.	Measurement model of 135 economics and accounting teachers' conceptions of assessment	72
Figure 4.2.	Distribution of responses from items contributing to factor – assessment describes learning.....	85
Figure 4.3.	Distribution of responses from items contributing to factor – assessment has weaknesses.....	86
Figure 4.4.	Distribution of responses from items contributing to factor – assessment improves learning and teaching.	87
Figure 4.5.	Distribution of responses from items contributing to factor – assessment for school accountability.....	88
Figure 5.1.	Overview of themes derived from the qualitative data.....	100

Terms and Abbreviations

Māori	Indigenous New Zealander.
NCEA	National Certificate of Educational Achievement: New Zealand's secondary school qualification.
NZC	New Zealand Curriculum
NZCETA	New Zealand Commerce and Economics Teachers' Association.
NZQA	New Zealand Qualifications Authority, the government organisation that administers NCEA.
OECD	Organisation for Economic Cooperation and Development.
Pākehā	Māori language word for New Zealanders who are of European descent. A term that is commonly used to describe this demographic grouping in New Zealand.
PPTA	New Zealand Post Primary Teachers' Association (union that represents New Zealand secondary school teachers).
TKI	A Ministry of Education bilingual web-based community that provides educational material for teachers, including exemplars for NCEA internal assessment.

CHAPTER 1

Introduction and background to the study

The National Certificate of Educational Achievement (NCEA) is New Zealand's main national qualification for senior secondary school students. NCEA is a standards-based assessment system which was introduced progressively between 2002 and 2004. It replaced the former norm-referenced qualification system of School Certificate, University Entrance and Bursary (New Zealand Qualifications Authority; NZQA, n.d.a), and was introduced because the previous system was perceived to be inadequate for assessing the education that young people require in a changing world (Barker, 1995). NCEA specifically sought to provide more opportunities for young people to succeed and achieve qualifications, as well as promote a culture of life-long learning (NZQA & Organisation for Economic Co-operation; OECD, 2004). This disruption to "entrenched practices and hierarchies" (Hipkins, Johnston, & Sheehan, 2016, p. 4) was not without controversy, and implementation of a standards-based system for school qualifications, such as NCEA, has not been common internationally (Shulruf, Hattie, & Tumen, 2010). A key feature of NCEA is that it allows for more than 50% of students' final grades to be assessed by teachers during the school year through school-based assessment, known in New Zealand as *internal assessment*.

Teachers working with NCEA are of interest to study because although they arguably work in a qualifications-oriented education culture, they also have, by international standards, a comparatively substantial role in decisions regarding the format and content of assessment for qualifications. Despite this responsibility, little is known about teachers' practices in relation to internal assessment and what might be the influences on those practices. There has been some research conducted on the practices of teachers in relation to specific internally-assessed standards within NCEA, for example, science investigations (Hume & Coll, 2009; Moeed, 2010) and foreign language spoken proficiencies (East, 2016), but none with teachers of accounting and economics. These disciplines are of interest, and were chosen as a lens for investigating conceptions of assessment and internal assessment practices, because

they are taught predominantly in the senior secondary school. The main – and often the sole – assessment focus for teachers of these subjects is on national qualifications.

1.1 The structure of NCEA

NCEA is considered to be one of the most complex school qualification systems in the world (Hipkins et al., 2016) and schools and teachers have considerable flexibility in how they choose to implement it. This section provides a brief introduction to the structure of NCEA and how students achieve NCEA qualifications.

NCEA is predicated on elements of assessment called standards that specify outcomes and criteria that students must meet to gain credits towards qualifications (Hipkins et al., 2016). Students usually work towards the three levels of NCEA in their last three years of schooling with Level 1 generally being attempted in Year 11, Level 2 in Year 12, and Level 3 in Year 13 (NZQA, n.d.b).

There are two types of standards that can contribute credits towards NCEA: achievement standards and unit standards. Achievement standards assess learning goals from the New Zealand Curriculum (NZC), whereas unit standards are more often linked to vocational training and are predominantly, but not exclusively, used by polytechnics and industry training organisations (NZQA, n.d.c). A further difference between achievement and unit standards is that most unit standards¹ operate on a pass/fail basis (rather than awarding a range of grades), whereas achievement standards specify criteria for three levels of achievement – Achieved, Merit, Excellence – and also have a Not Achieved/fail grade. All unit standards are internally assessed (i.e., assessed by teachers during the year), whereas some achievement standards are internally assessed and others externally assessed, generally by an end-of-year examination. Each subject is limited to three externally-assessed achievement standards at each of the three levels of NCEA.

¹ A limited number of unit standards do have the different grade levels available.

The assessment of a school subject for NCEA comprises a number of standards, with each assessing an area of content or skill from the subject domain. Students accumulate credits from standards towards NCEA qualifications at one of its three levels. Standards carry varying numbers of credits with those in accounting and economics ranging from three to six credits. In addition, to gain any level of NCEA students must also meet a minimum level of literacy and numeracy, which can be gained from either prescribed achievement standards or specific unit standards (NZQA, 2011). The requirements to gain NCEA at each level are outlined in Table 1.1.

Table 1.1
Requirements to gain NCEA at each level

NCEA level	Requirements
Level 1	80 credits from either level 1, 2 or 3, including literacy and numeracy.
Level 2	60 credits at level 2 or above, plus 20 credits from any level. Level 1 literacy and numeracy requirement.
Level 3	60 credits at level 3 or above, plus 20 credits from level 2 or above. Level 1 literacy and numeracy requirement.

The number of credits available in each subject domain, and at each NCEA level, varies. Table 1.2 shows that in economics and accounting the numbers of available credits range from 24 to 30.

Table 1.2
Total number of standards and credits available at each level for accounting and economics.

	Level 1	Level 2	Level 3
Accounting	27 15 internal/12 external	26 14 internal/12 external	26 13 internal/13 external
Economics	24 12 internal/12 external	30 18 internal/12 external	24 10 internal/14 external

If a student studies five subjects at Level 1, then an average of 16 credits per subject is needed to achieve NCEA at that level, and if they have six subjects (which is common at Year 11), then an average of only 14 credits per subject is required. So for levels 2 and 3, if a student has credits from the previous level then they only require an additional 60 credits, an average of 12 credits per subject. As students do not need all available credits in each subject to gain NCEA, teachers are able to alter their assessment programmes (and possibly their curriculum) to offer fewer than the total number of available standards. Students themselves can also choose which standards to attempt, and choices can be made as to the number of externally-assessed versus internally-assessed standards that will be attempted.

This flexibility and choice can make the system responsive to students' needs, and they can be assessed on a large variety of tasks and activities (Education Review Office, 2007; Hipkins et al., 2016). Schools can use the standards to assess courses based on traditional subject areas and assess them using combinations of internally- and externally-assessed achievement standards, as well as unit standards (if appropriate). Alternatively, standards from different subjects can be grouped to assess cross-disciplinary courses. For example, NZQA (n.d.d) cites Alfriston College designing a course called 'People, Places and Events', which is assessed using a combination of Level 1 History, Geography and Social Studies achievement standards. A further flexibility is that standards from different year levels can be attempted within the same course.

On the other hand, there can be drawbacks to the modular approach to assessment because course design can be fragmented and become aggregations of topics and standards rather than being based on a coherent curriculum (Hipkins et al., 2016). There is also a possibility that teachers (and students) only value material that will be assessed and that teachers will only teach that content (Pedulla et al., 2003; Stobart, 2008). To date, there is little research about what teachers are doing with respect to their use of the flexibility that NCEA affords, particularly in relation to teachers of economics and accounting.

NCEA certificates and courses can be endorsed with Excellence or Merit to recognise higher levels of aggregate achievement. To gain an Excellence endorsement for an NCEA certificate a student must gain 50 credits or more at Excellence in the level of the certificate or higher. Excellence for a course (subject) is gained by achieving 14 or more credits at Merit or Excellence, with at least three of these credits from externally-assessed standards, and three credits from internally-assessed standards (NZQA, n.d.e). A layer of complexity is that teachers may also need to ensure that students are able to gain university entrance at Level 3. The requirements for university entrance are:

- NCEA Level 3.
- Three subjects at Level 3, made up of 14 credits each, in three approved subjects (of which economics and accounting are both included).
- Literacy – 10 credits at Level 2 or above.
- Numeracy – 10 credits at Level 1 or above.

A further adjunct to NCEA is New Zealand Scholarship, which is an award made to the top achievers in secondary schools. Scholarship is examined largely by external examinations which are separate to NCEA examinations and are undertaken by relatively few students (Hipkins et al., 2016). Standing completely separately from New Zealand qualifications, but offered in some New Zealand schools in addition to NCEA, are the International Baccalaureate and the Cambridge International Examinations. These are further considerations for teachers to factor into their programmes. NCEA could be described as one of the most flexible, but complex assessment systems in the world and these additional qualifications further complicate assessment in the senior secondary school. This thesis focuses on NCEA, but teachers in New Zealand schools are potentially dealing with a more complex combination of qualifications and assessment demands.

According to NZQA (2017a) NCEA devolves much greater responsibility for course design and assessment to schools and teachers than the previous system did, because internal assessment accounts for approximately 70% of the students' results annually. The ability to choose which assessment standards will be used for a course, in addition to the internally-assessed component, gives schools opportunities to develop specific

programmes, to assess students in authentic ways, and provides students with opportunities to show what they can achieve in familiar surroundings and with ample time (Education Review Office, 2007; NZQA, n.d.f). However, teachers have to balance this flexibility with ensuring that students can meet the requirements for each NCEA level as well as course and certificate endorsements, and university entrance.

1.2 Economics and accounting in NCEA

Economics and accounting are long-standing subjects in New Zealand secondary schools; both were taught prior to NCEA and were examined under the former School Certificate, Sixth Form Certificate, and Bursary system. Both subjects were included in the NZC Framework (1993) but only economics was supported by a Ministry of Education curriculum document; it was up to the New Zealand Commerce and Economics Teachers' Association (NZCETA) to develop such a document for accounting (NZCETA, 2006). The NZC (Ministry of Education, 2007) omitted accounting, much to the dismay of accounting teachers who argued strongly for it to be included in the Social Science learning area (NZCETA). Despite this omission, achievement standards and unit standards have continually been provided, and supported, for both subjects since the inception of NCEA.

The popularity of economics and accounting has been waning for some time, with the numbers of students studying economics at all levels falling from 35,000 in 2003 to 24,000 in 2015 and those in accounting falling from 21,000 in 2003 to 15,000 in 2015 (Education Counts, n.d.). A new subject, Business Studies, which has been developed since the advent of NCEA, may have gained some of these students with 21,000 enrolled for 2015 (Education Counts). These issues are relevant because it is generally economics and accounting teachers who teach Business Studies, and some teachers add Business Studies achievement standards to their accounting and economics programmes.

The structure of NCEA economics and accounting is summarised in Tables 1.3 and 1.4. The numbers of credits available at each year level exceed those needed to gain a course in NCEA, so teachers and students make choices regarding which standards

to include in assessment programmes. In addition, the numbers of credits available for each course, at each level, from internal assessment range from 10 to 15, which also allows for a degree of choice, and provides scope for teachers to create programmes to meet students' needs and interests.

Table 1.3

Economics matrix for NCEA (Ministry of Education, n.d.a)

Level 1	Level 2	Level 3
AS90983 4 credits. External Demonstrate understanding of consumer choices, using scarcity and/or demand.	AS91222 4 credits. External Analyse inflation using economic concepts and models.	AS91399 4 credits. External Demonstrate understanding of the efficiency of market equilibrium
AS90984 5 credits. Internal Demonstrate understanding of decisions a producer makes about production.	AS91223 4 credits. External Analyse international trade using economic concepts and models.	AS91400 4 credits. External Demonstrate understanding of the efficiency of different market structures using marginal analysis.
AS90985 3 credits. External Demonstrate understanding of producer choices using supply.	AS91224 4 credits. External Analyse economic growth using economic concepts and models.	AS91401 5 credits. Internal Demonstrate understanding of micro-economic concepts.
AS90986 5 credits. External Demonstrate understanding of how consumer, producer and/or government choices affect society, using market equilibrium	AS91225 4 credits. Internal Analyse unemployment using economic concepts and models.	AS91402 5 credits. Internal Demonstrate understanding of government interventions to correct market failures.
AS90987 4 credits. Internal Demonstrate understanding of a government choice where affected groups have different viewpoints.	AS91226 4 credits. Internal Analyse statistical data relating to two contemporary economic issues.	AS91403 6 credits. External Demonstrate understanding of macro-economic influences on the New Zealand economy.
AS90988 3 credits Internal Demonstrate understanding of the interdependence of sectors of the New Zealand economy.	AS91227 6 credits. Internal Analyse how government policies and contemporary economic issues interact.	
	AS91228 4 credits Internal Analyse a contemporary economic issue of special interest using economic concepts and models.	

Table 1.4
Accounting matrix for NCEA (Ministry of Education, n.d.b)

Level 1	Level 2	Level 3
AS90976 3 credits. External Demonstrate understanding of accounting concepts for small entities.	AS9117 4 credits. External Demonstrate understanding of accounting concepts for an entity that operates accounting subsystems.	AS91404 4 credits. External Demonstrate understanding of accounting concepts for a New Zealand reporting entity.
AS90977 5 credits. Internal Process financial transactions for a small entity.	AS91175 4 credits. Internal Demonstrate understanding of accounting processing using accounting software.	AS91405 4 credits. Internal Demonstrate understanding of accounting for partnerships.
AS9097 5 credits. External Prepare financial statements for sole proprietors.	AS91176 5 credits. External Prepare financial information for an entity that operates accounting subsystems.	AS91406 5 credits. External Demonstrate understanding of company financial statement preparation.
AS90979 4 credits. Internal Prepare financial information for a community organisation's annual general meeting.	AS91177 4 credits. External Interpret accounting information for entities that operate accounting subsystems.	AS91407 5 credits. Internal Prepare a report for an external user that interprets the annual report of a New Zealand reporting entity.
AS90980 4 credits. External Interpret accounting information for sole proprietors.	AS91481 4 credits. Internal Demonstrate understanding of a contemporary accounting issue for decision-making.	AS91408 4 credits. External Demonstrate understanding of management accounting to inform decision-making.
AS90981 3 credits. Internal Make a financial decision for an individual or group.	AS91179 3 credits. Internal Demonstrate understanding of an accounts receivable subsystem for an entity.	AS91409 credits. Internal Demonstrate understanding of a job cost subsystem for an entity.
AS90981 3 credits. Internal Make a financial decision for an individual or group.	AS91386 3 credits. Internal Demonstrate understanding of an inventory subsystem for an entity.	

1.3 Rationale for the study

The internally-assessed achievement standards potentially provide scope for teachers to make decisions about the content to be assessed and the conditions under which the assessment will take place. There is also scope at each level to vary the curriculum (by varying the standards to be assessed) for individual students or groups of students. In their study on teachers' assessment practices for Level 3 NCEA, Stewart, Gray, and Pilcher (2007) found that teachers varied the numbers of standards they offered, and

allowed students to choose the standards on which they would be assessed. They also found that some teachers were designing their own assessment activities and/or modifying Te Kete Ipurangi (TKI)² tasks for the internally-assessed standards. However, other research tends to suggest that such innovative practices are not particularly widespread. Moeed (2010) reported that Year 11 science teachers tended to use the same TKI internal assessment resource for science achievement standard 90186 (Carry out a practical science investigation, with direction), and had done so for the previous eight years. Hipkins (2010) found that only 3% of teachers in her survey agreed that they involved students in co-creating NCEA plans, while 64% said that this never or almost never occurred. Hipkins also found that most principals and teachers agreed or strongly agreed that a range of assessment methods can be valid for NCEA, but her findings prompted her to ask if “teachers simply agree in principle, or are they actively extending the scope of assignments in practice” (p. 28). This question could equally well have applied to Wylie and Bonne’s (2016) finding that nearly 70% of teachers in their survey agreed that NCEA *allows* them to design courses to meet students’ learning needs, but without any evidence that this actually occurs. Edwards (2013) suggests not, claiming that many teachers do not develop their own assessment tasks; instead, they use TKI exemplars or commercial materials. Alison (2005) made similar assertions; however, there is still much to find out about how teachers are assessing the internally-assessed achievement standards and the motivation behind their choices. Are they giving students one-off, time-limited test-type activities, or, as suggested by NZQA and Education Review Office, providing students with ample time to show what they can do? Are they being creative and innovative in their design of assessment activities and the place of those activities within their course? It is an aim of this research to contribute answers towards these questions, in particular for teachers of economics and accounting.

Research into teachers’ beliefs (Borko & Putman, 1996; Fives & Buehl, 2016) suggests that beliefs are a major influence on what teachers do in the classroom and subsequently their assessment practice. G.T.L. Brown (2008) refers to beliefs about assessment as conceptions and, because of teachers’ extensive experience of assessment, they have

² TKI is a Ministry of Education bilingual web community that provides educational material for teachers, including exemplars for NCEA internal assessment.

developed conceptions of assessment and there is evidence that teachers' conceptions of assessment will mediate how they implement it (Bonner, 2016; Richardson & Placier, 2001; Thompson, 1992). It might, therefore, be predicted that teachers' conceptions of assessment will have an impact on how they implement the internally-assessed achievement standards. Additionally, conceptions of assessment are not fixed and can be influenced by external factors (e.g., G.T.L. Brown, 2008; Brown, Hui, Yu, & Kennedy, 2011; Brown, Kennedy, Fok, Chan, & Yu, 2009; Harris & Brown, 2009), so assessing for NCEA may have an impact on those conceptions. However, there is little existing evidence regarding the impact of NCEA on teachers' conceptions of assessment and little on how their conceptions influence their implementation of the internal assessment component of NCEA.

1.4 The researcher's position

I am a former teacher of accounting and economics, now working in initial teacher education. My current position includes teaching prospective economics and accounting teachers, which made me aware of the lack of research on this particular group of teachers. As a former secondary school teacher of economics and accounting for 25 years, I have always had a strong interest in assessment for qualifications. For most of those 25 years, the secondary school sector underwent assessment reform that generally saw a move towards senior secondary school qualifications including internally assessed components. In the 1980s, Sixth Form Certificate was introduced as a Year 12 qualification that was entirely internally assessed. These assessments were statistically moderated to a pre-determined national distribution (NZQA, n.d.g). This period also saw the introduction of internal assessment components for two-thirds of School Certificate and Bursary subjects, including economics, but not accounting.

The 1990s saw the introduction of fully internally-assessed unit standards for all subjects; School Certificate, Sixth Form Certificate, and Bursary were continued alongside the new assessment system. The school in which I taught at the time fully adopted the new unit standard regime. There were three strong drivers for this adoption: the capacity for unit standard assessment to be integrated with learning; the ability for students to receive immediate feedback on reaching the standards; and for students to

achieve credits towards a qualification progressively throughout the year rather than relying totally on an end-of-year examination. My involvement extended beyond school; I was appointed to NZQA expert panels to write and rewrite unit standards. I also served as a moderator for unit standards and, subsequently, internally-assessed achievement standards for NCEA.

In 2000, it was announced that NCEA would be introduced in 2002 and that all assessment for qualifications would use internally- or externally-assessed achievement standards or unit standards. At this point NZQA called for educators to conduct and facilitate workshops for the implementation of the new system. Based on my previous experience with standards-based assessment, I was appointed as a facilitator to deliver professional development for economics teachers, particularly with regard to internally-assessed achievement standards.

Upon appointment to a College of Education teaching position, I continued my interest in NCEA internal assessment through participation in two research projects evaluating NCEA professional development (Starkey, et al., 2006; Taylor, Kinsella, Yates, McKenzie, & Meyer, 2007). My contact with teachers through NCEA professional development, research, and visiting student teachers on their teaching placements resulted in exposure to widely varying opinions on NCEA. I was surprised at the negativity (sometimes vehement) expressed towards NCEA, which was not solely a personal impression; Alison (2005) reported that about two-thirds of teachers were either opposed to or ambivalent about NCEA. Teachers' support for NCEA has improved over the years, and Wylie and Bonne (2016) reported that in 2015 nearly 70% of teachers they surveyed either agreed or strongly agreed that they support NCEA. Even so, this leaves 30% not supporting NCEA. My own view had been (and still is) supportive of NCEA and its goal of improving student achievement in the senior secondary school by allowing for the assessment of a broader range of skills and knowledge.

As a researcher, I am declaring my longstanding interests favouring internal assessment and am aware of the potential for my own perspectives to bias my interpretation of the data. However, as a researcher I am committed to a primary and overarching responsibility to collect valid and reliable data, to carry out accurate analyses of the data, and interpret findings consistent with the evidence.

1.5 Thesis overview

In this opening chapter, I outlined the rationale for the study and explained the structure of NCEA, in particular the structure of accounting and economics within NCEA. Chapter 2 provides a review of the literature relevant to the purposes of assessment and how teachers use assessment, literature on the impact of assessment policy on teachers' beliefs and practices, and of the literature examining teachers' conceptions of assessment and teachers' assessment identity. Chapter 3 outlines the research design and the rationale for using mixed methods. It also addresses how the data were analysed and the inference process chosen to converge the two strands of the data. Chapter 4 presents the findings from the quantitative data and chapter 5 the findings from the qualitative interviews. Chapter 6 discusses the quantitative and qualitative findings contiguously in order to answer the first two research questions and chapter 7 considers how both strands of data connect and converge. Chapter 8 concludes the thesis by outlining the key contributions of this thesis, the limitations and implications.

CHAPTER 2

Literature Review

2.1 Defining assessment

Assessment is fundamental to education because it provides a means of finding out what students have learned (Wiliam, 2008). Gipps (1994), Harlen (1994) and Wiliam (1994) have emphasised that assessment is a way of gathering evidence that can be used for decision-making to inform future learning, whereas Taras (2010) asserted that educational assessment essentially entails a judgement and that the outcome of the judgement can be represented by a grade, by marks, or by feedback. Assessment has been conceptualised as a process contributing to improving learning (the formative use of assessment) and in contrast, as a strategy to measure what has been learned (the summative use of assessment).

In their seminal literature review on assessment and classroom learning, Black and Wiliam (1998) described the formative function of assessment as assessment that takes place during the learning process to inform students, and their teachers, of what will be required for students to take their next learning steps. To enable this, Scriven (1966) emphasised the need to look beyond grades and percentages as indicators of learning and encouraged teachers to analyse students' work to ascertain where any lack of understanding has occurred; that is, understanding the nature of students' mistakes and omissions is important if assessment is to improve learning. Sadler (1989) agreed that the focus of the formative use of assessment is to recognise the gap between what students currently know and what they need to know, and that feedback is a key element in this regard. Agreeing with this position, and moving towards a more precise meaning of the formative function, Black, Harrison, Lee, Marshall, and Wiliam (2002) further refined the description by adding that formative assessment requires that the evidence gathered during an assessment process to be used to adapt teaching to meet learning needs. In addition, Harlen (2006) considered that students should be active participants in the formative process; not only should the teacher act

on the feedback, but students should also decide on possible actions based on the feedback. Taras (2010) added a stricter proviso that assessment can only be considered formative “when feedback has been used to improve the work” (p. 3021). Taken together, the formative function of assessment includes making judgements on students’ learning in order to ascertain the gap between actual learning and desired learning, using that information to decide a course of action, which may include the teacher changing their pedagogy and should include the student in the process. Ultimately, the formative use of assessment should lead to improved learning.

The summative function of assessment, on the other hand, is to estimate what students know and can do at any point in time. It is usually described by a grade and is often geared towards certification (Sadler, 1989). The summative purpose of assessment is an integral part of the secondary school system and can have positive outcomes for students in the form of qualifications attainment (Madaus & Russell, 2011). It also provides a record of students’ achievements for themselves, their parents, future employers, and higher education institutions (Harlen, 2006).

There has been debate on the usefulness of the summative function of assessment to the learning process with Black and Wiliam (1998) claiming that the summative function of assessment was not particularly useful to future learning. Sadler (1989) agreed that summative assessment does not normally improve learning, but it often influences decisions and “may have profound educational and personal consequences for the student” (p. 120). In what appears to be a modification of Black and Wiliam’s (1998) earlier position, Black, Harrison, Hodgen, Marshall, and Serret (2011) used empirical evidence from a qualitative longitudinal study with 18 teachers to describe how working towards the summative process can help students’ learning. They concluded that, if students are actively involved then they are able to see how summative functions of assessment, for example tests and examinations, might help their learning. Further to this argument, and building on an earlier stance, Black, Harrison, Lee, Marshall, & Wiliam (2003) considered that the summative function of assessment could and should be a “positive part of the learning process” (p. 56).

Despite the distinction between these two functions being seemingly clear, and the terms *formative* and *summative* designed to clarify the purposes of assessment, Newton (2007) has argued they have in fact caused confusion. Confusion about the functions and purposes of assessment has possibly arisen because educationalists attributed intrinsic properties of the assessment process to the terms *formative* and *summative*. These terms refer to the purposes of the assessment rather than to the timing or the nature of the tasks. Black and Wiliam (1998) argued that the two purposes were at opposite ends of a spectrum, that they are “two isolated and completely different functions” (p. 35). Taras (2010) agreed that the terms formative and summative have led to confusion between the *process* and *purpose* of assessment and argues that assessment is a “single process with more than one function” (p. 3018). She added that the distinction between the formative and summative functions of assessment is that for a formative process to occur, the feedback should lead to improved learning, so any assessment without feedback and improved learning is summative. Furthermore, she argued that formal work used for qualifications, along with students’ work produced in less formal situations, can perform a formative function as long as learning is improved. Despite the confusion caused by the terminology, Harlen (2006) has suggested that the dual functions of assessment are reason enough to maintain the distinction.

2.2 School-based (internal) assessment

Assessment can be put to more than one purpose: it can be used for a formative purpose, a summative purpose, or for both purposes. There is also a possibility that it may have no purpose at all. Rather than considering the summative and formative functions of assessment as being polarised, Harlen (2006) has suggested that there should be more focus on how potential synergies between the two functions could be developed. While there is some concern that if teachers have to provide summative information about students, this can undermine formative assessment purposes (Nusche, Laveault, MacBeath, & Santiago, 2012), internal assessment for NCEA lends itself well to being used for both purposes and possible synergies could be realised.

Internally-assessed standards for NCEA are those that are assessed by teachers during the year and are sometimes used to assess skills and knowledge that cannot be tested in end-of-year of examinations (NZQA, n.d.f). Internal assessment involves teachers in all stages of the assessment cycle; they plan the assessment programme, design or choose assessment tasks, and make judgements about the final grade. In addition, internal assessment allows students to have more than one attempt to achieve standards, meaning that earlier attempts can provide formative information to support later attempts.

Internal assessment lends itself to both formative and summative functions because it provides opportunities for feedback and learning to occur, while at the same time allowing teachers to make summative judgements on what students know and can do at any point in time. Rawlins (2008) argued that NCEA internally-assessed standards have the potential “to satisfy both the summative and formative purposes of assessment” (p. 115) because they meet Sadler’s (1989) three characteristics needed for assessment to be considered formative. They have criteria that can be shared with students so that they know what they are aiming for, feedback can be provided as students work towards achieving the criteria, and they can use the teachers’ feedback and self-assessment to recognise their next learning steps. Of course, to realise a formative purpose, students must engage with the feedback and improve their learning. As noted above, NCEA internally-assessed standards explicitly provide this opportunity by allowing students further opportunities to reach the standard.

The Ministry of Education (2011) concurs with Rawlins’ (2008) view and stated that even though NCEA standards constitute high-stakes assessment, this should not preclude them from being used to support learning. The Ministry of Education stated that the design of internal assessment for NCEA “provides the potential for assessment to be used formatively and to be an integral part of teaching practice” (p. 14). NCEA internally-assessed standards allow synergies between the formative and summative functions of assessment to be developed with Black et al. (2011) arguing that teachers’ internal assessments can promote formative practices because teachers are in control of the assessment process. Furthermore, their empirical data suggested that students often become more involved, and actively engaged, in a summative process when it is managed by teachers.

Internal assessment allows for a much greater range of tasks to be used than externally-assessed examinations; course work can include a range of assessment activities including projects, research, essays, presentations, portfolios and group work and “course work lends itself to formative assessment practice, as teachers provide feedback and appropriate guidance to students” (Ministry of Education, 2011, p. 69). From a sociocultural perspective, the flexibility (in terms of tasks, the timing of assessment, and resit opportunities) afforded by internal assessment is beneficial because, under this view, assessment activities should be used to engage students in learning and “to increase their participation and sense of belonging” (Klenowski & Wyatt-Smith, 2014, p. 33). Klenowski and Wyatt-Smith proposed that inclusive assessment practices can involve teachers adapting assessment tasks to meet the diverse needs of students more effectively and this can be achieved through internal assessment.

Nonetheless, tensions can arise from attempts to use assessment for both formative and summative purposes. In particular, fairness and authenticity are important considerations in assessment for qualifications, but measures taken to assure authenticity can limit the extent to which assessment can be effective in a formative sense. A time-limited test is a case in point; the time limit and the uniformity of the task deliver a high probability of equal assessment conditions and minimise opportunities for cheating, but also place constraints on the format of the assessment task itself and so may limit its formative value. Harlen (2006) agreed that there are limitations to using data gathered primarily for one purpose in another way, noting that data gathered for summative purposes may lack the detail required to address specific learning needs. Similarly, evidence gathered for formative purposes may be inconsistent because it relates to the specific learning goals of individual students and is bound by the context of their learning. Harlen proposed that both functions of assessment can be served from the same evidence as long as a distinction “is made between the evidence and the interpretation of the evidence” (p. 109). Carless (2011) agreed that school-based assessment has the potential to combine the formative and summative functions of assessment.

For the formative use, assessment data will inform what further learning needs to be done, while at the same time the data can fulfil a summative purpose by describing the level that a student has reached at that point in time. For NCEA internal assessment teachers could ascertain from students' classwork whether they have met criteria, for example, the criteria required to achieve the standard at the Achieved level. They could provide feedback to improve the work to an Excellence or Merit level and at some point in time decide which achievement level the student has reached.

Maxwell (2004) offered possible synergies that can be obtained between the two functions of assessment and suggested that portfolios of evidence gathered over time can serve both formative and summative functions. Countering the argument that assessment should be the same for all students in order for it to be fair in summative terms, Maxwell contended that evidence towards an assessment does not have to be identical to be comparable. He argued that the contents of student portfolios (for example) could differ, but still be judged against common standards. In addition, portfolios or other types of assessment gathered over a period of time can involve feedback to students and therefore perform a formative, as well as a summative, function.

Maxwell (2004) and Harlen (2006) provided theoretical possibilities for the formative and summative functions of assessment to be implemented within internal assessment; however, empirical studies have found that teachers vary in their implementation of internal assessment in practice. Yung's (2012) qualitative case study of 10 Hong Kong science teachers reported different approaches by teachers to practical science assessment. Some teachers conducted these separately from classroom practice, in test-like conditions, which generates few opportunities to provide formative feedback, whereas others made assessment judgements in the course of the students' learning, while at the same time giving them formative feedback. While Yung's conclusions were drawn from a relatively small group of teachers, survey research from larger groups have reported similar findings. Reporting on survey data from 152 language teachers in New Zealand, East (2016) also found that teachers' reactions to internal assessment varied; some conducted internal assessment in the course of learning programmes and provided feedback as students worked on assessments, whereas others conceived of the internal assessment as a one-off assessment event, such as a

test, with little feedback provided. Also, in New Zealand, Moeed (2010) reported on a survey of 189 science teachers who described varying practices. Some carried out practical science assessments in test-like conditions, with 78% stating that they carried out formative preparation as a trial run or ‘mock exam’ of the actual internal assessment. In their in-depth qualitative study of 11 New Zealand secondary school teachers Irving, Harris, and Peterson (2011) reported that teachers were reluctant to use internal assessment results for a formative purpose when there was also a summative requirement. Despite expressing a preference for using internal assessment in a formative way, they noted tensions between this pedagogical aim and requirements to report results for summative purposes. A reason given for this dilemma was that teachers in the study were not clear about which practices could be considered formative and they based the distinction between summative and formative functions on whether or not a grade was given.

Internal assessment affords the possibility of gaining synergies between the formative and summative functions of assessment and authors such as Maxwell (2004), Harlen (2006) and Rawlins (2008) have made suggestions as to how this could occur. However, achieving such synergies with NCEA internal assessment has proven difficult to achieve in practice (East, 2016; Irving et al., 2011; Moeed, 2010).

2.2.1 Issues of validity and reliability

If teachers’ judgements are used to award students national qualifications, their assessment tasks must be valid and their marking reliable to ensure fairness for all students. According to the American Educational Research Association, American Psychological Association and the National Council on Measurement in Education (2014), validity is “the degree to which evidence and theory support the interpretation of test scores for proposed uses of tests” (p. 11) which is reminiscent of Messick’s (1975) definition that validity requires that assessment must measure what it is intended to measure and that it supports the inferences that are drawn from the assessment.

Validity has been differentiated into types which include (among others): content validity; predictive validity; and construct validity. Content validity is the extent to which an assessment measures against curriculum coverage; predictive validity is the extent to which an assessment can be used to predict some future outcome; and construct validity refers to the measurement of an underlying psychological quality (American Psychological Association, American Educational Research Association, & National Council on Measurement in Education, 1954; Newton & Shaw, 2014). Not all academics have agreed with these fragmented definitions and Messick (1989) rationalised that construct validity incorporated the ideas of content and predictive validity, but added that valid assessment should also take into account the ethical consequences of assessment. Newton and Shaw provided a comprehensive history of the concept of validity in educational assessment, but argued that the term is still “heavily contested” (p. 1). In proposing a twenty-first century framework for validity, Newton and Shaw, suggested that for assessment to be valid it must be both technically possible and socially acceptable, and that this may involve “trading costs against benefits” (p. 187). Therefore, the argument is not so much about whether one type of assessment is more valid than another, but rather whether an assessment policy as a whole is acceptable, which Newton and Shaw acknowledge may differ “from one stakeholder to the next” (p. 188). What is acceptable in one assessment jurisdiction may not be acceptable for another and in New Zealand, this would be acceptability of the structure of the NCEA qualification that includes internal, teacher-assessed standards as well as externally-examined standards.

Achieving validity can be complex and questions have been raised about teachers’ ability to construct valid tasks for internal assessment. Davison and Leung (2009) claimed that there is no guarantee that teacher-made assessment tasks assess what is intended, and Johnson (2011), while acknowledging that there is little research on teacher-conducted internal assessment, argued that “different teachers might operationalise their assessment of a subject domain differently” (p. 3). In her systematic literature review, which concerned the validity of teachers’ assessment for summative purposes, Harlen (2004) defended teacher-conducted assessments, arguing that because teachers build a picture of student learning over a range of objectives, they are able to provide a fuller account of student achievement than can be obtained through tests.

Reliability refers to the degree to which an assessment process produces consistent and stable results over markers, time and tasks (Brookhart, 2003; Harlen, 2007) and marker reliability has been raised as a concern for internal assessment. Johnson (2013) argued that there is little empirical evidence for the reliability of teacher judgements, calling for more research to be done in this area. She provided some examples from Australia and Scotland to exemplify the issue. In Australia, the Queensland Studies Authority reported 87% marker agreement and only serious disagreement in 4% of cases, suggesting a good level of reliability. However, the Scottish example revealed lower agreement levels of 45% for numeracy, around 40% agreement for reading, and between 10% and 35% for science, figures that question the reliability of the marking of the teachers concerned. Hay and McDonald (2008) in their study of Australian physical education teachers reported that, although teachers were provided with criteria and standards, they would make judgements about students' levels of achievement without reference to these. The teachers believed they had internalised the criteria, but Hay and McDonald reported that the judgements were intuitive and reliant on the teachers' memories of students' past performances.

Evidence from New Zealand suggests that marker reliability can improve over time. NZQA (2016) reported that agreement rates between teachers and moderators for grade level in NCEA had improved from 79.6% in 2012 to 84.6% in 2015. Arguments around agreement rates need to consider the scale of measurement; it would be more difficult to get agreement if the mark range is 0-100 than if the range is A-E, or in the case of NCEA achievement standards which afford only four possible grades. Evidence from longitudinal studies has supported the notion that reliability of teacher judgements can improve and become more consistent with professional development. In their two-and-a-half year qualitative collaboration with 18 teachers, Black et al. (2011) reported that despite initial variability in teacher judgements, with ongoing professional development the consistency of judgements improved over time. Similarly, Adie, Klenowski, and Wyatt-Smith (2012), also reporting on a longitudinal study, which investigated how teachers came to shared judgements about student work, concluded that the reliability and consistency of teachers' judgements did increase over time.

2.2.2 Validation and moderation

Systems can be put in place to ameliorate deficits in the validity of assessment tasks and reliability of teachers' marking. Johnson (2011) described 'controlled assessment' for the General Certificate of Secondary Education in the United Kingdom, whereby the Qualifications and Curriculum Development Agency provided a range of tasks that teachers had to select from, rather than allowing them to create their own. The agency provided marking schemes, and prescribed levels of supervision under which the assessment had to take place. The system operating in New Zealand for assessment material is more akin to support than control; NZQA provides exemplar materials for NCEA for teachers to adapt and use if they choose. Control is asserted over teachers' marking because the validity of a proportion of tasks set by teachers is checked by moderators (NZQA, n.d.h). While these measures may alleviate validity concerns, they are in tension with some possible purposes of internal assessment. For example, Klenowski and Wyatt-Smith (2014) and Maxwell (2004) suggested that internal assessment allows for assessments to be personalised to meet the needs of a diverse range of students, which is not possible if all students are expected to complete identical assessments. Davison and Leung (2009), in their article promoting a theory for internal assessment, attempted to explain the validity dilemma by arguing that a fundamental paradox of internal assessment is that its strengths can also be considered weaknesses. They claim that the validity of internal assessment is derived from teachers knowing students' work and not from similarity of tasks.

Moderation is a practice that involves supporting the reliability of teachers' judgements. Linn (1993) describes two key ways of carrying this out: statistical moderation and social moderation. Statistical moderation involves students' marks in one type of assessment being checked, and sometimes altered, on the basis of a statistical comparison with another assessment (Linn); for example, internal assessments may be moderated against an external examination (Hipkins et al., 2016). Social moderation involves groups of teachers sharing their understandings of standards, and students' work towards the standard, in order to develop common judgements (Adie et al., 2012; Hipkins et al., 2016; Linn, 1993). According to Linn, social moderation also includes teachers' judgements being rated independently by teachers external to the school or other independent experts. This latter description of

moderation is referred to as external check moderation in NCEA and the external expert is known as a moderator (NZQA, n.d.h).

Nevertheless, such strategies to achieve greater reliability have been criticised. Statistical moderation addresses anomalies in marking, but Johnson (2013) questions the value of carrying out internal assessment if statistical moderation is then applied. She also points out that a weakness of check moderation is that the moderation takes place after the assessment and “any injustices seen to be done to individual students by their teachers’ assessments remain unidentified, or, if identified, unaddressed” (p. 99). A further weakness is that moderators become custodians of standards and, according to Johnson, the validity and reliability of their judgements are open to question. A major weakness of social moderation is that consistency among teachers cannot be addressed if there is no external system to moderate across the groups (Hipkins et al., 2016). A further weakness reported by Adie et al. (2012) was that changing teachers’ judgements through social moderation is a complex process, influenced by many factors, but as argued by Linn (1993), “staff development is a critical component” (p. 99). Arguing against the possibility of total reliability, Davison and Leung (2009) suggested that internal assessment does not require teachers to give identical marks; some variation within a range should be acceptable.

Assessment jurisdictions that include some form of internal assessment must consider the reliability of teachers’ marking and the validity of their assessments to ensure fairness for all students. However, concern about these is not, in itself, a reason for not including internal assessment in national qualifications. Systems such as moderation can be put in place to ameliorate potential deficits in this regard. The validity of teachers’ internal assessment tasks and judgements is derived from teachers knowing their students’ work and not from the similarity of assessment tasks. As Newton and Shaw (2014) have argued, for assessment to be ethical and socially acceptable, the benefits of one type of assessment often have to be traded against those of another.

2.2.3 Impact of internal assessment on teachers

Internal assessment can support positive changes in teachers' practice, such as: promoting the provision of formative feedback during the assessment process (Harlen, 2004); heightening their awareness of their teaching effectiveness (Meyer, McClure, Walkey, McKenzie, & Weir, 2007); and, enabling them to use a wider range of assessment tools with the autonomy to design programmes to meet the needs of their students (New Zealand Post Primary Teachers' Association; PPTA, 2015). Further benefits for teachers of using internal assessment are having their professionalism acknowledged, empowering them by recognising their extensive knowledge of their students as learners and developing high levels of assessment skill (Johnson, 2013). Evidence from New Zealand has shown that teachers do believe that they are professionally supported in their internal assessment practice. In a national survey of secondary teachers, Hipkins (2013) reported that 75% of teachers attending Best Practice Workshops (professional development offered by NZQA for internal assessment) said these had increased their confidence in making assessment decisions. Hipkins also reported that teachers thought moderation procedures provided them with useful feedback to improve their assessment skills. In her research on internal assessment systems in the United Kingdom, Queensland and California, Harlen (2007) similarly claimed that feedback from moderation can provide valuable professional development and improve teachers' assessment practices.

In contrast, Taylor et al. (2007) reported that teachers interpreted feedback on their marking and assessments as questioning their professionalism, a sentiment echoed by the PPTA (2016). The PPTA, in their investigation into teacher workload issues, which consisted of a survey of 1,300 secondary teachers and focus group interviews, reported that teachers thought that their ability to assess accurately was questioned by the current levels of moderation required by NZQA, levels that participants in the PPTA research considered excessive. However, NZQA (2017a) claimed that moderation requirements are relatively low and operate as a high-trust model because only approximately 4% of results reported are required for moderation. While it is in the interests of teacher unions to make claims in favour of teachers, independent studies have similarly reported that teachers can find moderation processes critical, inconsistent, and unhelpful professionally (Brooks, 2010; East, 2016; Mizutani, 2009;

Wylie & Bonne, 2016). Contrasting opinions on teachers' perceptions of their professionalism could be a result of differing levels of competency with regards to internal assessment. Those showing higher levels of competency have their skills endorsed by external procedures, while those less competent or confident in internal assessment may think their professionalism is more exposed and therefore report more negative feelings. Nonetheless, the most important question is perhaps whether or not the reliability of teachers' marking improves as a result of the process.

Teachers in a number of jurisdictions have raised concerns about the workload associated with carrying out internal assessment. In their review of practices of internal assessment in Australia, Scotland and Wales, Stanley, MacCann, Gardner, Reynold, and Wild (2009) identified teacher workload as an issue, and called for internal assessment systems to be manageable for teachers and not to require over-assessment or excessive record keeping. Workload issues have also been reported in Hong Kong, New Zealand, and Malaysia. For example, in Hong Kong, Yip and Cheung (2005), reporting on a quantitative study of 351 teachers, noted that teachers were very concerned about the high workload associated with internal assessment with many regarding it as additional work imposed by examination authorities. Yan's (2014) quantitative study with 280 participants similarly reported that teachers considered carrying out internal assessment to be a burden and workload concerns contributed to teachers' negative attitudes towards internal assessment. In a qualitative study, Qian (2014) added that the heavy workload of internal assessment may result in teachers not producing quality work and that excessive workload can have negative effects on teachers' well-being.

In New Zealand, teachers and senior leaders in schools have criticised the increased workload created by internal assessment and moderation requirements (East, 2016; Mizutani, 2009; PPTA, 2016; Wylie & Bonne, 2016). Wylie and Bonne, in the triennial survey of secondary teachers run by the New Zealand Council for Educational Research, reported that the impact of internal assessment on teacher workloads "has been an issue since its inception" (p. 26). Only 32% of teacher participants in their survey agreed that the workload was manageable, with over 40% disagreeing. An even greater proportion of principals (84%) agreed that teachers struggled with the workload created by internal assessment.

Studies that include interviews as well as surveys allow participants to expand on the sources of workload pressure. According to the PPTA (2016), internal assessment was the “most referenced source of workload pressure” (p. 68) with moderation requirements being cited as a source of considerable workload. East (2016) reported that language teachers participating in his study found it time-consuming and complicated to carry out internal assessment because multiple assessment opportunities were being offered for students to meet spoken language proficiency standards. Further sources of workload that have been reported are a lack of resources and support for developing new assessment materials, inconsistent moderator feedback, and an increase in administration associated with moderation and record keeping for internal assessment (Mizutani, 2009).

In their major review of the Queensland system of senior assessment, which received 2,200 survey responses, Matters and Masters (2014) referenced workload issues associated with internal assessment and argued that one advantage of introducing some external assessment back into the Queensland system would be a reduction in teachers’ workload. Mills and McGregor (2016), in their analysis of the Queensland education system, agreed that a challenging aspect of internal assessment is the “great increase in teacher workload” (p. 114).

Three studies from Malaysia have also reported workload issues related to implementing internal assessment. Majid (2011), in a quantitative study investigating teachers’ concerns with assessment innovation, reported that teachers felt that they did not have enough time to implement internal assessment and that too much time was spent on administrative tasks associated with it. Furthermore, time pressure led to negative feelings about internal assessment. Although Majid reported on a relatively small ($n=40$) group of teachers, her findings are supported by a larger study that surveyed 322 teachers and also concluded that the heavy workload associated with internal assessment leads to negative attitudes towards it (Jaba, 2013). Raman and Yamat (2014), in their qualitative study similarly identified workload concerns, with participants revealing that the sources of extra workload were the marking, filing, documentation and paper work associated with internal assessment.

A consideration with regard to workload concerns is that, if teachers saw the formative value in internal assessment they may have a different view. If they considered internal assessment a valuable part of the learning process that at the same time awards credits towards national qualifications rather than an administrative imposition, their attitudes towards the workload could be different.

2.2.4 Impact of internal assessment on students

Harlen (2004) reported on an in-depth literature review on the impact of using internal assessment for summative purposes. At the time, only six studies were exclusively concerned with the use of internal assessment for senior secondary school qualifications. From this limited database, Harlen concluded that internal assessment had a positive effect on students: students often find internal assessments motivating; they value managing their own learning; and that they appreciate becoming more independent in terms of assessment. Empirical studies since then have reiterated such positive outcomes of internal assessment for students. Sheehan (2013), in his mixed methods study of New Zealand history students, concluded that through internal assessment students were developing understandings of how history works as a discipline and were learning how to think historically. He claimed that the internally-assessed course work was making a major contribution to these students “learning to participate in society as critical citizens who can think independently and adjudicate between competing claims of historical authenticity” (p. 11). Furthermore, he reported that students in his study were motivated to engage with internally-assessed course work, that they valued the autonomy of investigating issues of personal interest, and that they were prepared to commit to the substantial workload required. Also using mixed methods to report on the internal assessment component of NCEA, Mizutani (2009) identified several benefits for students. These included: more authentic assessment, for example, conversations in a target language as opposed to written translations; students being more aware of expectations through specific assessment criteria; ongoing formative feedback; and students taking more control of their own learning and feeling motivated to try to achieve higher grades.

Also, taking a mixed methods approach in their study of the impact of NCEA on student motivation, Meyer et al. (2007) reported positive perceptions of the internally-assessed component of NCEA. They found that internal assessment provides achievement opportunities for students who may be unsuccessful in examinations as well as increasing the choice of topics to study. It allows assessment workload to be spread over an entire year, and can afford opportunities to resubmit or repeat assessments. Student participants in Mizutani's (2009) study also reported that they appreciated having assessments spread out over a year because that gave them opportunities for feedback, which assisted their learning by giving them clear goals to work towards.

Some concern has been expressed about student workload with regard to internal assessment, for example, by Wylie and Bonne (2016). However, this concern has often emanated from parents and teachers rather than the students themselves. For example, Mizutani (2009), who surveyed and interviewed students, reported that while teachers thought ongoing assessment throughout the year was a burden to students, the students themselves did not agree. Students reported that they found the system motivating and manageable.

The international literature on student voice in internal assessment is rather sparse. Two studies from Hong Kong have reported varying student attitudes towards internal assessment. Gao's (2009) small-scale quantitative study showed that students within one school had contrasting views, with some thinking the internal assessment process was highly beneficial, while others thought the feedback from teachers was inadequate. Tong and Adamson (2015) followed up Gao with a larger mixed methods study of 423 survey respondents and interviews with 45 students that reported mixed responses: over 70% of respondents did not agree that internal assessment led to a less stressful assessment experience, but some participants did think that the feedback was supportive of learning. Overall, they concluded that students did not appreciate internal assessment and that they were dissatisfied with the feedback they received from teachers, considering it insubstantial and imprecise.

In these examples, it seems the utility of internal assessment to students is mediated by its implementation by teachers. Mizutani (2009) reported that students in her study found that teachers' feedback assisted their learning, and that they were generally positive about internal assessment. On the other hand, students in the studies of Gao (2009) and Tong and Adamson (2015) were disappointed with teacher feedback and were subsequently not supportive of internal assessment.

2.2.5 Concluding comment

Using internal assessment for external qualifications has the potential to benefit students and teachers alike. Students can be assessed on a greater range of skills that cannot feasibly be tested under examination conditions and assessment can be carried out in less stressful conditions; thus it is likely to give students more opportunities to demonstrate their full potential. Issues of validity and reliability of teacher assessment have to be considered and systems put in place to ensure fairness. In an environment with internal assessment for external qualifications, teachers must balance the multiple purposes of assessment and the workload associated with carrying out a substantial level of summative assessment. We need to know how they react to this challenge to ensure that change is implemented as intended.

2.3 Assessment policy as a tool to improve teaching and learning

2.3.1 Assessment reform in New Zealand

Assessment is not an isolated educational phenomenon; it is essentially a social tool used to judge performance (for formative and summative purposes) and often regulates access to further opportunity. Barker (1995) argued that education and assessment policy were increasingly being used to promote social, political and economic agendas. As a result of New Zealand's weakening economy in the 1970s and 1980s there were calls for changes in the education and qualifications systems as these were now perceived to be inadequate in educating and assessing people to work in a changing economy. Numerous reports (e.g., *Education in Change*, PPTA, 1969; *Towards partnership: The report of the Committee on Secondary Education*, 1976; *Learning and achieving: Second report of the commission of inquiry into curriculum, assessment and qualifications in forms 5 to 7*, 1986; *Assessment for better learning*, 1989, *Assessment: Policy to practice*, 1994) advocated such change, arguing that the

existing normative assessment model, which ensured compulsory failure of approximately 50% of students, could not support a nation of successful achievers (Barker, 1995).

Very little change took place until 1990 when NZQA was established under the 1990 Education Amendment Act, with the main purpose of reforming qualifications, including those assessed for secondary school students. The New Zealand Qualifications Framework (NQF) was introduced in 1991 (NZQA, 1991). Its features included standards-based assessment that would be assessed by unit standards. NZQA introduced unit standards progressively from 1992 with an intention to completely replace the norm-referenced examination-based system with internally-assessed unit standards by 1996 (NZQA, 1992). The school assessment system was to move simultaneously to a standards-based regime that would be fully internally assessed.

The move to standards-based assessment was criticised by a number of education academics. Elley (1995) claimed that the use of standards-based assessment for high-stakes assessment is suspect and argued that setting specific standards for some skills such as writing fluently or critically examining texts was not possible. He further argued that the standards proposed by NZQA were not precise enough and terms used to describe criteria within standards “like ‘some’, ‘an inclusive range’ and a ‘reasonable amount’ are inexplicit and unhelpful” (p. 83). These views were echoed by Crombie (1995) who asserted similarly that standards could not be used to assess higher-level skills in languages. She believed that only low-level skills could be assessed using criteria-based standards. While not criticising standards-based assessment *per se*, Hall (1998) raised serious issues with the structure of the unit standard system that modularised the assessment for each subject into small components. He was also wary of the proposal for total reliance on internally-assessed standards and supported a system that used both internal and external assessment. Hall claimed that clearly defining performance standards was difficult and the application of standards “inevitably involves some element of norm-referenced assessment” (p. 38), because the proportions of students in large cohorts gaining various grades is not expected to vary statistically from year to year (Hipkins et al., 2016). He argued that rather than dismissing either norm-referenced or standards-based assessment due to their weaknesses, the strengths of each approach should be used in tandem.

In the meantime, NZQA continued to offer two systems: the NQF with its unit standards and the traditional norm-referenced examinations of School Certificate and University Bursary. Schools and individual students were able to choose between these systems, or in some cases, operate in both. Research (e.g., Ministry of Education, 1996) with schools using unit standards found that teachers were generally positive about the potential of the NQF and unit standards to deliver a fairer qualification structure for their students. Some positive aspects they noted were that the system was flexible, that reassessment was possible, and that it allowed more students to gain qualifications.

In 1997, the Ministry of Education released a draft policy document: *A Future Qualifications Policy for New Zealand: A Plan for the National Qualifications Framework*. The paper acknowledged the debate surrounding the NQF and proposed a way forward to develop a cohesive qualifications system. The draft policy proposed to retain unit standards as many providers had “enthusiastically embraced the unit standards-based NQF” (p. 15), but to incorporate the existing school examinations into the NQF as their supporters “believe they identify excellence in student performance better” (p. 26). *Achievement 2001*, the policy for school qualifications, was announced in November 1998 with the main points of the policy being:

- All students would study towards NCEA.
- In conventional curriculum subjects, credits would be gained from a mix of internal and external assessment that would be against achievement standards.
- Achievement standards would recognise merit and excellence.
- Unit standards would remain. (Baker, 2001)

At the time of its launch, NZQA claimed that NCEA would “provide a more comprehensive record of what students achieve while they are at school. It would also give students a better launching pad for ongoing learning and their future careers than the previous system” (NZQA, n.d.a, p. 3). The progressive introduction of the NCEA between 2002 and 2004 moved secondary school assessment from a norm-referenced paradigm to a standards-based one, with an additional intention of recognising, and giving credit for, a wider range of knowledge and skills than the previous school

qualification system (NZQA, n.d.a). According to Meyer et al. (2007), NCEA was intended to achieve particular educational and social outcomes such as higher proportions of students staying in school in Years 12 and 13, and higher qualifications attainment. However, resistance to NCEA remained and led to some secondary schools adopting foreign qualifications such as the Cambridge International Examinations and the International Baccalaureate (Fitzpatrick & Locke, 2008).

2.3.2 The impact of assessment reform

The implementation of NCEA in New Zealand secondary schools has been described as a ‘seismic’ change in assessment (Hipkins et al., 2016) and the launch of the new qualifications system in 2002 was followed by the introduction of a new school curriculum in 2007. These policy reforms reflected high expectations for student achievement and anticipated a significant shift in secondary teachers’ practice, requiring them to change their thinking about assessment and their classroom practice (Hipkins et al.). Hargreaves (1989) warned that assessment reforms would not automatically result in teachers changing their pedagogy with which Torrance (1995) agreed and added that expecting a straightforward relationship between assessment reform and improved teaching and learning is unrealistic. Such a relationship suggests that politicians make policy and teachers “simply do as they are told and put the intended policy into practice” (Thrupp, 2017, p. 12). In reality, policy is reinterpreted and implemented by teachers within the complexities of diverse school settings, along with other education policies and wider social issues (Ball, Maguire, & Braun, 2012).

Fullan (2001) explained that education reform is complex because if teachers are to implement change as it is intended, they need the motivation to do so, and this may require a change in their beliefs. However, if they think that their core beliefs are challenged and contradicted, teachers may be resistant and very little actual change will take place. Furthermore, teachers may need to change their pedagogical practices and learn new skills, which creates further barriers to change, even if teachers endorse the reform. Finally, education reform takes time; teachers must add new knowledge to existing schema requiring sufficient resources to support any change taking place (Noble & Smith, 1994).

The challenge of implementing assessment and curriculum reform in New Zealand rested with secondary classroom teachers, because, as Senge (1990) claimed, the success of any reform hinges on what happens at the smallest unit of the organisation. For educational reform this is the classroom in which teachers were challenged to implement and embed change. Despite the major change to school-level qualifications, there is little empirical evidence regarding secondary teachers' beliefs about, and practices within, the current assessment regime, in particular its internally-assessed component.

On the other hand, the impact of assessment reform has been measured through publically reported assessment data. Percentages of students attaining each level of NCEA are gathered and published, and in 2012 the State Services Commission (New Zealand Government, 2012) announced a goal which declared that by 2017 85% of all 18-year olds would achieve NCEA Level 2. In addition, according to Hipkins et al. (2016), some secondary schools have set assessment achievement targets for teachers in response to the aforementioned goal. Broadfoot (2007) argues that using assessment data that are gathered for one legitimate purpose for another, arguably a less legitimate one, is an unfortunate function of assessment activities. Broadfoot provides the example of assessment data designed to measure the performance of individual students then being used to evaluate schools and subsequently to create league tables comparing schools and teachers, with penalties applied to those that do not show favourable data. Broadfoot contends this is the least helpful and most damaging use of assessment results because any evaluation of schools and teachers should be rigorous and carried out with consideration of its consequences. This point of view has long been argued. For example, in their historical overview of accountability and evaluation in education Madaus, Scriven, and Stufflebeam (1983) agreed that test scores should not be a sole measure of school accountability. That it is still being argued in 2016 (e.g., Nichols & Harris, 2016) indicates that such use of assessment results still occurs despite the practice being manifestly unfair because it does not account for external factors that may affect student, and therefore school, performance.

Broadfoot (2007) argues that using assessment data to evaluate schools is associated with performativity with performativity referring to a culture that employs judgements, comparisons, rewards and punishments as means of incentive and control. Broadfoot claims performativity is not compatible with the aims of standards-based and internal assessment, both of which aim to set clear learning goals for students and to assess a wide range of skills. Conversely, she claims, performativity encourages extrinsic motivation in students, is socially divisive, and appears to raise standards while at the same time driving out assessment for empowerment and learning. Broadfoot quotes the children's writer David Almond lamenting this trend:

There is an arrogance at work, the arrogance that we know exactly what happens when someone learns something, that we can plan for it, that we can describe it, that we can record it and if we can't do these things, then the learning doesn't exist. (Almond, cited in Broadfoot, 2007, p. 68)

Using assessment policy as a tool for educational reform creates dilemmas. In New Zealand, assessment reform in recent decades occurred against a background of unprecedented change in the education system and wider society, which some authors have argued was driven by a New Right agenda (Codd, McAlpine, & Poskitt, 1995). The stated purposes of the assessment reform were to give students a better grounding for future learning, to provide a more comprehensive record of achievement, and to recognise, and give credit for, a greater range of knowledge and skills than the previous system (NZQA, n.d.a). However, data from those assessments are also used to evaluate schools' and teachers' performance, a practice that can create underlying tensions, contradictory purposes, and confusion for teachers (Bonner, 2016). Bonner considers that there is still work to do to understand the impact of major assessment change on teachers' perceptions of, and actions towards, assessment.

2.4 Teachers' beliefs

Beliefs are judgements regarding the validity of truth claims (Pajares, 1992) and according to Bandura (1997), they guide our decisions and actions. Beliefs can facilitate or hinder practice by serving to filter, frame and guide practice (Fives & Buehl, 2016). According to Borko and Putman (1996), teachers' beliefs are a major influence on what they do in the classroom. If policy change is to influence teachers' practice, then attention needs to be paid to the beliefs they currently hold, so that teachers can consider those beliefs and evaluate how they facilitate or hinder their practice (Fives & Buehl, 2016).

A variety of terminology has been used to define beliefs, for example, *cognition* (Borg, 2003), *perceptions* (Bonner, 2016), *values* (James & Pedder, 2006). Thompson (1992) suggested the use of the term *conceptions* because a conception is "a more general mental structure [than a belief], encompassing beliefs, meanings, concepts, propositions, rules, mental images, preferences, and the like" (p. 130). G.T.L. Brown (2002) also preferred this term because "conceptions represent different categories of ideas" and act as a framework "through which a teacher views, interprets, and interacts, with the teaching environment" (p. 2). Conceptions can be erroneous, contradictory and incomplete, but they do represent what we understand and believe, and they guide our actions and attitudes (G.T.L. Brown, 2008). In addition, conceptions of assessment are complex and suggest a dynamic relationship between beliefs and classroom practice in that beliefs guide actions, but alternatively actions can influence beliefs (Looney, Cumming, van Der Kleij, & Harris, 2017).

Teachers hold conflicting conceptions of assessment that vary with context and the developmental position of students (Bonner, 2016). G.T.L. Brown and his colleagues have studied the underlying structure of teachers' beliefs about assessment and have developed a model that describes four predominant conceptions (e.g., G.T.L. Brown, 2004; Brown et al., 2011; Brown et al., 2009; Brown, Lake, & Matters, 2011; Brown & Remesal, 2012; Gebril & Brown, 2013).

G.T.L. Brown's (2002) doctoral dissertation used a series of qualitative and quantitative studies to test models of how teachers' conceptions of assessment were structured. He found that the views New Zealand teachers held about assessment aligned with a theoretical model comprising four broad categories. These are that: assessment is for the *improvement* of teaching and learning; assessment is for making *schools accountable* for their effectiveness; assessment is for making *students accountable* for their learning; and that assessment has no purpose and is *irrelevant* or detrimental to the life and work of teachers and students.

The *improvement* conception of assessment asserts that the purpose of assessment is to generate valid information that will lead to improved student learning, which may entail changes in teaching and/or student learning practices. As noted by G.T.L. Brown (2008), this conception assumes that unless improvement in learning takes place, assessment could be conceived as being irrelevant. This proposition aligns strongly with Wiliam's (1994) assertion that there is little point in carrying out assessment if there is no improvement to student learning and therefore the *improvement* conception encapsulates the formative function of assessment.

According to G.T.L. Brown (2008), the *school accountability* conception of assessment relates to the belief that assessment results can provide surety to the government that its resources are being used well in schools and that the mandated curriculum is being taught. In addition, the *school accountability* conception assumes that assessment data can make schools and teachers accountable for their work, and assessment results can lead to improved teaching and thereby raise educational standards. If teachers use assessment data to analyse their own practice, this can potentially lead to improvement in students' learning; therefore, according to Brown (2008), a positive correlation between this conception and the *improvement* conception could be expected. On the other hand, if the consequences of accountability are too high (e.g., schools vilified or teachers reprimanded for poor results) then this may lead to teachers considering assessment antithetical to their welfare and to conceiving of this function of assessment as irrelevant (G.T.L. Brown, 2008).

The *student accountability* conception holds students responsible for their own learning through the assignment of grades and, in the senior secondary school, through the award of various qualifications (G.T.L. Brown, 2008). Grades signal that students have, or have not, attained a standard for entry into the next levels of education or employment, thus holding students accountable for having achieved a certain level of competence. This aspect of the *student accountability* conception would align with a summative purpose of assessment (G.T.L. Brown, 2008). According to G.T.L. Brown (2008), teachers attribute accountability to students because their effort and attention in class are seen to contribute to their learning. A further practice that may contribute to teachers conceiving of assessment as an indicator of *student accountability* is the increasing role some teachers give to students to determine their own learning goals (G.T.L. Brown, 2008). For example, in a secondary school this could include allowing students flexibility to choose which NCEA achievement standards they will attempt. Interviews with teachers could determine if this is the case and whether such practices are associated with student accountability.

There is also a conception that assessment has no legitimate place in education beyond the intuitive processes that teachers engage in with students. Under this conception, formal assessment is seen as *irrelevant*. This belief stems from the ideas that assessment has an unfair impact on some students, that assessment can be so inaccurate that it has no reliability, and that assessment is anchored in a “non student-centred view of the educational world” (G.T.L. Brown, 2008, p. 29).

2.4.1 Measuring teachers’ conceptions of assessment

G.T.L. Brown (2002) developed an instrument comprising an inventory of 50 statements to measure the dimensionality of teachers’ conceptions of assessment and the degree to which teachers endorse each of the four conceptions. G.T.L. Brown used exploratory and confirmatory factor analysis over a series of studies to develop a model of how teachers’ conceptions of assessment are structured. He concluded that the data gathered using the instrument “were consistent with the theoretical models on which they were based” (p.186). Subsequently, G.T.L. Brown (2006) reported on the effectiveness of an abridged version of only 27 items – the Teachers’ Conception of Assessment Abridged (TCoAIIIA) Inventory.

Quantitative research using the TCoAIIIA Inventory undertaken in New Zealand and Australia has supported the model (G.T.L. Brown, 2004, 2006, 2011a; Brown, Lake, & Matters, 2011). However, one key difference between primary and secondary teachers in these countries was noted. In Australia, Brown, Lake, and Matters reported that primary teachers agreed strongly with the *improvement* conception whereas secondary teachers had higher agreement with *student accountability*, findings which agreed with similar New Zealand research (G.T.L. Brown, 2011a). Despite this difference, G.T.L. Brown (2011a) and Brown, Lake, and Matters (2011) concluded that there was a strong similarity amongst the responses of primary and secondary teachers in New Zealand and Australia. Similarities included positive correlations between the *improvement* and *student accountability* conceptions, an inverse correlation between the *irrelevant* and *improvement* conceptions, and a positive correlation between the *irrelevant* and *student accountability* conceptions. G.T.L. Brown (2011a) explained these correlations as showing that primary and secondary teachers agreed that if assessment does not improve learning then it can be considered irrelevant. The positive correlation between *irrelevant* and *student accountability* indicated a predominant belief that making students accountable for their own learning is seen to be an irrelevant function of assessment. G.T.L. Brown (2011a) attributed the similarity between New Zealand primary and secondary teachers to common professional attitudes endorsing a child-centred curriculum and concluded that the qualifications system “did not greatly distort teachers’ conceptions of assessment relative to primary teacher responses” (p. 63). This degree of commonality could be considered surprising given the strong focus on qualifications in the senior secondary school years. Further investigation, for example, with interviews, could explain more fully the choices secondary teachers made on the survey.

Quantitative studies with secondary teachers in more examination-oriented education systems have generally not supported the factor structure of G.T.L. Brown’s (2006) model, for example, in Hong Kong (Brown et al., 2009), Egypt (Gebril & Brown, 2013), southern China (Brown et al., 2011), and the Netherlands (Segers & Tillema, 2011). Brown et al. (2009) found that, while secondary teachers in Hong Kong had similar conceptions of assessment reported to those of the New Zealand teachers identified by G.T.L. Brown (2011a), there were two notable differences. First, for Hong Kong teachers the correlation between the *student accountability* and

improvement conceptions was .91 whereas for New Zealand teachers it was only .23. This strong positive correlation has also been reported in southern China (Brown et al., 2011) and Egypt (Gebril & Brown, 2013). Gebril and Brown postulated that this correlation indicates that secondary teachers in Hong Kong, China and Egypt hold students accountable for their own improvement. However, there are many ways to interpret correlations, and positive correlations between these two conceptions could also be explained in terms of teachers considering indicators associated with *student accountability* (e.g., placing students in categories, assigning grades, and meeting qualifications) as *evidence* of improved learning. Alternatively, this correlation could be an indicator of a belief that students who take responsibility for their own learning improve more over time. Interestingly, in the Netherlands, Segers and Tillema, who used exploratory factor analysis, did not extract *student accountability* and *improvement* as two separate conceptions; they speculated that teachers in their study did not make a distinction between the formative and summative purposes of assessment. Without further investigation, for example by asking participants, it is not possible to state categorically why these factor structures and correlations exist.

A further difference between the findings of G.T.L. Brown (2011a) and other studies is varying correlations with the *irrelevant* conception. Data from teachers in Hong Kong (Brown et al., 2009), China (Brown et al., 2011) and Egypt (Gebril & Brown, 2014) showed an inverse correlation between the *irrelevant* and *student accountability* conceptions, contrasting with a positive correlation for New Zealand teachers. This suggests that Hong Kong, Chinese and Egyptian teachers believe that assessment for student accountability has educational relevance, whereas New Zealand teachers are much more ambivalent in this regard.

2.4.2 Differences in teachers' conceptions of assessment

International differences in teachers' conceptions of assessment have been attributed to varying cultural expectations (Brown et al., 2009; Brown et al., 2011; Gebril & Brown, 2013; Segers & Tillema, 2011). The variations might also reflect the influence of national assessment policy on teachers' beliefs and, in turn, on their assessment practice. Alternatively, policy changes might affect practice first, leading to changes in belief, which would be in accordance with Bandura's (1986) theory of bi-directional influence. It is also possible that variations in teachers' conceptions of assessment are

related to the age level at which they teach (Remesal, 2011). Internationally, research has elicited variation in teachers' conceptions of assessment, with secondary teachers showing different conceptions from primary teachers (Brown, Lake, & Matters, 2011). Such differences could be influenced by the high-stakes nature of assessment for qualifications in secondary schools, or alternatively, by the nature of older students versus primary-aged students. Brown, Lake, & Matters, who used G.T.L. Brown's (2006) TCoAIIIA to compare Queensland primary and secondary teachers' conceptions of assessment, attributed differences between the two groups to high-stakes examinations in the senior secondary school which is "consistent with the notion that policy shapes conceptions" (p. 218).

Without asking teachers, researchers can only speculate on the reasons for their different conceptions of assessment. Research that includes interviews, as well as surveys, can provide such reasons because they afford participants an opportunity to elaborate on their survey responses. For example, in New Zealand, Brown and Harris (2009) carried out a mixed methods study, which reported that the introduction of a set of nationally provided resources to support teachers' use of assessment for learning moved teachers' beliefs from being predominantly improvement oriented to being more aligned with the idea that assessment measures school quality. Interviews were able to explain this unexpected result and revealed that some teachers could not reconcile the test-like nature of these resources with an improvement belief and that some schools did use the tool for their own, and their teachers', accountability.

A qualitative comparison between primary and secondary teachers in Spain (Remesal, 2011) showed that secondary teachers had a higher leaning towards *accrediting* conceptions of assessment (conceptions which focus on certification of achievement) than primary teachers. The qualitative nature of Remesal's study was able to reveal that the interviewed teachers hypothesised that their "conceptions could be related to the structure of the educational system and the external assessment policy demands" (p. 480).

G.T.L. Brown's (2011a) argument that teachers adopt conceptions of assessment that are "ecologically rational in that they reflect the legal, cultural or social priorities placed on assessment for their work environment" (p. 65) supports the hypothesis that teachers' conceptions of assessment may vary with the level at which they teach. Remesal (2011) reinforced this interpretation and, in addition, speculated that conceptions might differ between teachers of different subjects at the high-school level because her research focussed on mathematics teachers and therefore could be content specific. Mizutani, Rubie-Davies, Hattie, and Philp (2011) certainly found that the subjects teachers taught played a role in influencing beliefs about NCEA assessment.

Teachers' conceptions of assessment are complex – they vary according to context and the level at which they teach, and are far from fully understood. Bonner (2016) notes that the assessment policy landscape has changed substantially in the last decade and we do not fully understand the implications of this on teachers' conceptions of assessment. Outstanding questions include how and whether teachers' conceptions of assessment vary with the level at which they teach and whether the nature of the academic discipline taught by a teacher has any impact on his or her conceptions of assessment. Remesal (2011), on the basis of her findings and speculations, also called for more research investigating links between teachers' conceptions of assessment, their educational contexts, and the subjects they teach.

2.5 Teachers' assessment identity

Looney et al. (2017) claim that teachers' conceptions of assessment, and their experiences, feelings and beliefs about assessment are significant influences on how teachers assess. They "propose a dynamic and interactive teacher assessment identity constituted by beliefs, feelings, knowledge and skills" (p. 14). Their model of teacher identity also includes the impact of the role of the teacher in the assessment process. Xu and Brown (2016) also discuss the idea of teachers having an assessment identity and claim that teachers "need to constantly negotiate their role as assessors (p. 158), especially during periods of change. They also claim a challenge to teachers' assessment identity is the positioning of teachers as assessors of formal qualifications. Discussing a socio-cultural approach, Pryor and Crossouard (2008) claim that

teachers' identities "rather than being rooted in individual traits" (p. 9) are determined by socio-cultural and historical settings, because according to the socio-cultural approach people become who they are by participating in communities. Assessing students take place within the classroom community and the wider community of the school.

The notion of teacher assessment identity is still somewhat vague. There is a consensus in the literature that teachers have one, but little regarding what the elements of that identity might be, nor how identities might vary from one teacher to another. Xu and Brown (2016) claimed "there is no 'ideal' assessor identity" (p. 158) and that if teachers have a clear sense of who they are as assessors they are more likely to make justified decisions around their assessment practice and to engage in more self-reflection. Adie (2013) claimed that teachers themselves identify as a 'type' of assessor. She describes teachers who thought of themselves as either hard and nit-picky markers, or as easy and fair. While not especially discussing assessment identity, Horsley (2012) identifies three different teacher identities with regards students' success in the New Zealand Scholarship examinations; she describes teachers as either catalysts, mavericks or inhibitors to students' achievement.

Assessment identity can also be conferred by others. Foucault (1979) discussed the power of examinations to afford or deny opportunities to students, and their role in classifying students. Bourdieu (1991) claims that tests have power and Bachman and Purpura (2008) claimed that tests "serve as both door-openers and gatekeepers". Decisions based on assessments can either provide access to resources and opportunities (door-openers) or deny such access (gatekeepers). Shohamy (2001) invests that power with the testers who become powerful individuals able to give or deny access to further opportunity. The metaphors of gatekeeping and door-opening are therefore used to describe the individuals and their practices rather than just referring to the tests. Davison (2004) has also used metaphors to describe teachers' assessment practices; she refers to 'assessor-technicians' who followed procedures, rules and principles when assessing students and also 'assessors as God' who tended to make intuitive, subjective decisions about students' levels of achievement. While gatekeeping has negative connotations, Nagy (2000) argues that gatekeeping is a legitimate, historical role of assessment which acts a form of quality control.

With respect to NCEA internal assessment, teachers are the testers with the power to open doors or to act as gatekeepers.

2.6 Conclusion

Two decades of educational reform are a lived experience for many New Zealand secondary teachers. The impact of this reform, in particular the controversial assessment for qualifications reform, on their beliefs and practices is yet to be fully investigated. This review has outlined the educational importance of assessment and the tensions that can arise because assessment results are put to multiple purposes. Assessment reform aims to improve teaching and learning, as well as student achievement, but on the other hand, assessment results are used as an accountability measure for teachers and schools. Teachers have to manage these multiple purposes and intentions, which at times are contrasting and contradictory and the literature suggests ongoing tensions and conflict for teachers trying to balance competing functions of assessment. Teachers' beliefs have important implications for their actions, and in order to understand their actions, we must therefore attempt to understand their beliefs.

2.7 Research questions

Literature searches found a limited number of mixed methods studies in relation to teachers' conceptions of assessment, and none exclusively focussing on New Zealand secondary school teachers. In addition, there is currently little research, either quantitative or qualitative, specifically about New Zealand secondary teachers' conceptions of assessment. This group of teachers are of interest because they are working in a unique assessment environment in which they have a substantial role in decisions regarding the format and content of assessment for qualifications. This level of responsibility is likely to affect their conceptions of assessment, and their beliefs about assessment may determine their actions with regard to their assessment-for-qualifications programmes.

This thesis begins to address this gap by reporting the conceptions of assessment and practices towards internal assessment of a group of New Zealand economics and accounting teachers. A mixed methods approach was used to investigate the conceptions of assessment of these teachers, how they assess the internally-assessed achievement standards for NCEA, and the reasons for the decisions they make. Specific questions are:

- 1 What are the conceptions of assessment for teachers of accounting and economics?
- 2 What are the reasons for teachers of accounting and economics making the decisions they do about internal assessment tasks they offer students?
- 3 What is the impact of their beliefs about assessment on their approaches to internal assessment design?

2.8 Significance of the study

NCEA signified sweeping changes to assessment for qualifications in New Zealand secondary schools. It recognises and gives credit for a broader range of knowledge and skills than did the previous assessment system. The internally-assessed component, particularly, allows a wide range of student performance to be assessed and reported, which has enabled higher levels of student success in attaining school qualifications (NZQA, 2017b). As Yung (2000) suggested, “good school-based assessment can be a powerful tool for improving teaching and learning” (p. 267); therefore, it is important that internal assessment practices maximise this potential. Although New Zealand secondary teachers arguably work in a qualifications-oriented education culture, they also have a substantial role in decisions regarding the format and content of assessment for qualifications in comparison with teachers in many other nations. Furthermore, because a great deal of the assessment programme for NCEA is school-based, there is potential for this assessment to fulfil formative functions as well as contributing to awarding qualifications.

This study will build on, and contribute to, work previously done by G.T.L. Brown and his colleagues on teachers' conceptions of assessment (e.g. G.T.L. Brown, 2004; Brown et al., 2011; Brown et al., 2009; Brown, Lake, & Matters, 2011; Brown & Remesal, 2012; Gebril & Brown, 2013) by investigating a group of teachers not yet included in this research stream. The mixed methods design of this study adds a further dimension in investigating the reasons given by participants for their perspectives on, and conceptions of, assessment. This study contributes to ongoing research on NCEA and internal assessment by identifying the practices, and reasons for these practices, of a group of teachers towards the internally-assessed component.

This study is of significance to anyone (students, teachers, teacher educators, school managers, education officials) who is interested in realising the potential of internal assessment. The study will inform stakeholders of internal assessment practices of economics and accounting teachers in New Zealand schools. It will also highlight to stakeholders the enablers and barriers to teachers implementing 'good' internal assessment practice, which, in turn, can inform both initial teacher education and teacher professional development programmes. The findings of this research should be relevant to educationalists in other countries and regions wherever internal assessments inform the teaching and learning process as well as the awarding of secondary school qualifications.

CHAPTER 3

Methodology

3.1. Theoretical positioning of the study

According to Mackenzie and Knipe (2006), unless researchers nominate a research paradigm there is “no basis for subsequent choices regarding methodology, literature or research design” (p. 195). Morgan (2007) also claims that “research inherently involves epistemological issues about the nature of knowledge” (p. 52) and that the choice of research paradigm is shaped by the researcher’s preferences. The fundamental epistemic position underpinning the present research is a pragmatic world view. Pragmatism places methodology rather than philosophical stance at the centre of the research (Morgan) and allows the researcher to use whatever methods are appropriate to answer the research questions. Tashakkori and Teddlie (2003) argue that this is an appropriate approach for research in applied social sciences, and pragmatism offers an “alternative worldview to those of positivism/postpositivism and constructivism and focuses on the problem to be researched and the consequences of the research” (Feilzer, 2009, p. 7). Hence, using a pragmatic paradigm removes researchers from the post-positivist and constructivist dichotomy and allows them to see a world with both subjective and objective layers. As Dewey (1925, cited in Strauf, 2016) describes layers of “completeness, order, recurrences which make possible prediction and control, and singularities, ambiguities, uncertain possibilities, processes going on to consequences as yet indeterminate” (p. 47). The general focus of this research, NCEA internal assessment, is encompassed by Dewey’s words.

3.2 Research design

This research used a mixed methods design. According to Creswell and Plano Clark (2007), the mixed methods approach involves collecting and analysing both quantitative and qualitative data. It is appropriate to use mixed methods because such a design provides a more complete picture with the quantitative analysis uncovering

patterns and correlations, enriched by in-depth knowledge of perspectives provided by the qualitative analysis (Creswell & Plano Clark).

The research was designed to explore teachers' conceptions of assessment, their internal assessment practices, the reasons that these practices are occurring, and to investigate any relationships among them; therefore, one data source was insufficient. Teddlie and Tashakkori (2010) also claim that mixed methods research is valuable when the research problem is in a complex educational context. By using mixed methods it may be possible to achieve "a richer and more complete description of a phenomenon" (Willig & Stainton-Rogers, 2008, p. 14). To build a more comprehensive understanding of teachers' conceptions of assessment and the impact these have on their internal assessment practices, a survey exploring their conceptions of assessment was followed up with interviews. Willig and Stainton-Rogers (2008) warn that the results of the quantitative and qualitative elements of a mixed methods study may not necessarily converge, but note that they can nonetheless be used to critique each other.

Creswell and Plano Clark (2007) describe four major designs of mixed methods research: the triangulation design, the embedded design, the explanatory design, and the exploratory design. This research is most suited to the explanatory design because "the overall purpose of this design is that qualitative data help explain or build upon quantitative results" (Creswell & Plano Clark, p. 71). This design, also known as the Explanatory Sequential Design, is a two-phase design that starts with the collection and analysis of quantitative data which is followed by the qualitative data collection. This is useful if the researcher wants to use the quantitative results to form groups, or for purposive sampling, for the subsequent qualitative stage (Creswell & Plano Clark).

Creswell and Plano Clark (2007) describe this design as the most straightforward mixed methods design and one that can be carried out by a sole researcher because data are collected in separate phases and the research can be conducted as a single study. On the other hand, it may not be possible to specify how participants for the qualitative phase will be selected until the quantitative phase is complete. Furthermore, because of the inevitable delay between each of the phases, not all participants selected from questionnaires are likely to be available to take part in interviews.

3.3 Participants

3.3.1 Participant selection

Participants were teachers of accounting and economics teaching in courses assessed by NCEA internal achievement standards. These disciplines are taught predominantly in the senior school, so the main, and often the sole, assessment focus for teachers of these subjects is on national qualifications. Thus, participants are teachers with a predominant focus on assessing for qualifications, with responsibility for designing and implementing a substantial portion of that assessment, and with the opportunity to use that assessment for formative purposes.

3.3.2 Survey participants

One hundred and thirty-five teachers completed the conceptions of assessment part of the survey, and 119 completed the demographic section, although some did not answer all questions. Some key points about participants who shared this information are: 65% were female; most (80%) were experienced teachers with at least five years' experience; 98% had a graduate qualification; 86% were New Zealand trained teachers; 74% were New Zealand European; and all reported having some training in assessment. Half of the participants (50%) worked in state co-educational schools, but other types of schools were also represented including private schools (9%) and single-sex state schools (19%).

3.3.3 Interview participants

Initially, 55 respondents agreed to take part in interviews. However, due to a range of factors, for example, changing their minds when asked to take part, not being able to find a compatible date, sickness, and not responding to requests, only 21 were ultimately available to be interviewed. It was originally intended to purposively select interview participants based on different patterns of responses to the questionnaire but, because the numbers were manageable it was decided to interview all 21 who were available. The 21 interviewees represented a range of responses to the survey. Table 3.1 summarises some characteristics of each interviewee.

Table 3.1
Demographic characteristics of interviewees

Pseudonym	No. of years teaching	Gender	Ethnicity	School decile	School type
Angus	Over 20	Male	NZ European	High	State Single-sex
Scottiegh	Between 6 & 10	Male	European	High	State single-sex
Aisla	Between 11 & 20	Female	NZ European	High	Integrated single-sex
Barra	Over 20	Male	NZ European	High	Integrated single-sex
Catrina	Less than 2	Female	NZ European	High	State co-ed
Eilidh	Between 6 & 10	Female	Asian	Low	State co-ed
Mhairi	Over 20	Female	NZ European	High	State co-ed
Faibhe	Over 20	Male	NZ European	High	Private single-sex
Isla	Over 20	Female	NZ European	Low	State co-ed
Kirsty	Between 2 & 5	Female	NZ European	High	State co-ed
Ceanag	Between 6 & 10	Female	NZ European	High	State single-sex
Maeve	Between 11 & 20	Female	NZ European	High	State co-ed
Morag	Over 20	Female	NZ European	Low	State co-ed
Malcolm	Over 20	Male	NZ European	Low	State co-ed
Ruairidh	Between 2 & 5	Male	NZ European	High	Integrated single-sex
Sionn	Between 2 & 5	Male	Asian	Low	State co-ed
Seonaid	Between 11 & 20	Female	NZ European	High	Integrated single-sex
Sorcha	Over 20	Female	NZ European	High	Integrated single-sex
Tearlag	Between 11 & 20	Female	British	Medium	State co-ed
Wallace	Between 11 & 20	Male	NZ European	High	Private co-ed
Doileag	Between 11 & 20	Female	NZ European	High	State single-sex

3.3.4 Sample limitations

Table 3.2 summarises the demographic data of those who fully completed the survey and those who were interviewed, compared with national percentages of teachers.

Table 3.2

Percentages of participants in demographic categories compared with national figures.

Characteristics	Category	Survey %	Interview %	National %
Gender	Male	35	38	42
	Female	65	62	58
Teaching experience	Less than 2 years	7	4	Not available
	Between 2 & 10 yrs	26	43	
	Between 11 & 20	33	24	
	20 plus years	35	29	
Education level	Diploma	1	0	19
	Bachelors	30	33	Not available
	Graduate Diploma	15	4	Not available
	Post Grad Cert	7	9	Not available
	Post Grad Diploma	27	25	75
	Masters or higher	20	29	4
	None			2
Teaching subject	Economics	43	48	Not available
	Accounting	24	19	Not available
	Both	33	33	
Overseas trained	No	86	96	84
	Yes	14	4	16
Ethnicity	NZ European	77	81	79
	NZ Māori	2	0	8
	Pacific Nations	2	0	2
	Asian	7	14	3
	Other	12	5	7
Assessment training	None	0	0	Not available
	Pre-service	17	21	
	In-service	66	64	
	Undergraduate	6	0	
	Postgraduate	11	15	
Types of schools	State co-ed	50	48	56
	State single-sex	19	19	10
	Integrated secondary	23	19	22
	Private secondary	10	14	12
Decile rating	1–3 (low)	9	9	7
	4–7 (medium)	36	14	35
	8–10 (high)	55	76	58

Table 3.2 indicates that the sample of teachers who took part in the survey is fairly representative of the secondary teaching population, although there are some discrepancies. For example, nationally, 58% of secondary school teachers are female whereas 65% and 62% of participants in the survey and interviews, respectively, were female. The New Zealand Teacher Census (Ministry of Education, 2005) surveys

teachers by age, not years of experience, and links between age and years of experience can only be approximations, as older teachers may have started teaching later in life. However, if age can be considered a proxy for experience, the survey participants reflect the national averages fairly closely, although a preponderance of younger teachers (i.e., with 2 – 10 years' experience) took part in the interviews.

Ethnically, participants were distributed similarly to the national population with a large majority being New Zealand Pākehā³. There was similarity in terms of types of schools except that 19% of both survey and interview participants worked in state single-sex schools, whereas these types of schools represent 10% of the national total.

In terms of students studying economics and accounting, 7% are in low decile schools, 35% in medium decile schools and 58% in high decile schools (Education Counts, n.d). The survey participants reflect this distribution quite well, although proportionately more teachers from high decile schools agreed to be interviewed.

Although interviewees represented a range of experience, school type and location these were not completely reflective of the national population. As Katzer, Cook, and Crouch (1998) comment, “volunteers are different from those who do not volunteer” (p. 168) and one could speculate that teachers who are confident in their practice would be more willing to be interviewed than those who are not.

³ Pākehā is a Māori language word for New Zealanders who are of European descent.

3.4 Data collection methods

The following data sources were employed:

- self-report questionnaires.
- semi-structured interviews.

3.4.1 Self-report questionnaires

G.T.L. Brown's (2006) TCoAIIIA Inventory (Appendix 1) was used to investigate participants' conceptions of assessment. The TCoAIIIA Inventory elicits self-reported responses to 27 Likert-type items scored on a positively-packed, six point scale (*1 = strongly disagree, 2 = mostly disagree, 3 = slightly agree, 4 = moderately agree, 5 = mostly agree and 6 = strongly agree*). The items canvas agreement with a range of statements that align with four conceptions of assessment: *assessment for improvement; assessment for student accountability; assessment for school accountability; and assessment as irrelevant* (G.T.L. Brown, 2002).

G.T.L. Brown (2006) reported data validating the TCoAIIIA, an abridged form of G.T.L. Brown's (2004) 50-question Conceptions of Assessment-III. After studying data from 525 New Zealand and 692 Queensland primary teachers he concluded that the abridged version provides information of a similar quality to the long version, but is more efficient.

The TCoAIIIA has been used internationally as a research tool to examine teachers' conceptions of assessment, for example in New Zealand (G.T.L. Brown, 2011a; Harris & Brown, 2009), Hong Kong (Brown et al., 2009), China (Brown et al., 2011), the United States (Calveric, 2010), Cyprus (Brown & Michaelides, 2011), the Netherlands (Segers & Tillema, 2011), Australia (Brown, Lake, & Matters, 2012), Colombia (Muñoz, Palacio, & Escobar, 2012), Iran (Pishghadam, & Shayesteh, 2012), Spain (Brown & Remesal, 2012), Egypt (Gebriel & Brown, 2014), and Indonesia (Aziz, 2014).

In addition to the TCoAIIIA, the survey (Appendix 1) contained a number of questions eliciting demographic information, such as years of experience, gender, school decile, and the source and nature of the internal assessment tasks the teachers were using. The full survey was trialled with eight teachers who would not be participants in the study, and some adjustments were made to the layout and the wording of demographic questions based on feedback from this trial.

3.4.2 Survey administration and quantitative data collection

The survey comprised 51 questions and both paper and online versions were prepared. The online version used Qualtrics software and a forced-completion was used for the TCoAIIIA questions because the data would have been less valid if not all questions had been answered by all respondents. However, the demographic questions did not use a forced completion approach and 12% of participants chose not to answer these.

3.4.3 Survey qualitative questions

The survey contained a number of questions that invited a qualitative answer, for example, why teachers chose the methods and form of internal assessment tasks they use. These data were extracted from the quantitative data and coded separately.

3.4.4 Semi-structured interviews

Interviews were conducted to probe more deeply into the reasons that teachers make the decisions they do about their internal assessment tasks, the major influences on these decisions, whether they would prefer a different approach to internal assessment, and if so, why they do not enact this. The interviews also elicited how the participants defined and conceived of assessment in general. See Appendix 2 for the full generic interview schedule.

The interviews were face-to-face and semi-structured, so all participants answered the same or similar questions, but they also had opportunities to offer their own views and ideas. Semi-structured interviews allow the interviewer to go into more depth and resolve misunderstandings (Cohen, Manion, & Morrison, 2007) and give the researcher some control over the interview (Creswell, 2003). Face-to-face interviews give the interviewer the opportunity to judge the quality of responses, to give visual prompts such as nods and smiles, and to clarify if a question does not seem to have

been fully understood (Walliman, 2005). Open-ended questions were used because this type of question is flexible and can elicit unexpected answers, which may suggest relationships not previously considered by a researcher (Cohen et al., 2007). In terms of Patton's (2002) classification of types of interviews, the interviews most closely resembled the General Interview Guide Approach under which topics and issues are specified in advance but the interviewer decides on the sequence and wording of the questions as the interview proceeds.

All survey participants were invited to be interviewed and those who volunteered were contacted by email and asked to take part. If they agreed, they were sent further information about the project and the interview consent form. Subsequently, meeting times were arranged to fit with my travel arrangements and the interviewees' teaching commitments. All interviews were conducted face-to-face in the area in which the participants lived.

Participants were interviewed once, and with their permission all interviews were digitally recorded. These were subsequently transcribed verbatim, either by the researcher or by a professional transcriber, and returned to each interviewee for member checking. Member-checking allowed participants to correct, clarify or withdraw statements if they had changed their minds.

The general interview protocol included meeting and introducing myself, if necessary, or reacquainting ourselves. A number of the participants were known to me through economics and accounting teacher networks, or were former teaching colleagues. I thanked them for participating and agreeing to meet with me, reminded them of the general purpose and interest of my research, asked them to sign the consent form and then began with the open-ended questions. The duration of the interviews ranged between 45 and 90 minutes.

Interviews are not without limitations. According to Creswell (2003), interviews only provide "indirect information filtered through the views of the interviewees" (p. 187); the presence of the interviewer may bias responses and not all participants are equally articulate. Fontana and Frey (1994) warned of the impossibility of neutral interviewing, and my position as a former teaching colleague now undertaking a

doctoral dissertation may have elicited answers designed to impress, rather than participants' actual views and practices. Successful interviewing requires a researcher to build a relationship with an interviewee. Having been a teacher of economics and accounting in the NCEA assessment environment, I was able to establish rapport which enabled empathetic interviewing (Mertens, 2005). In order to mitigate some limitations of interviews, I used semi-structured questions so that all participants were asked the same or similar questions and all interviews were recorded so that the discussion could be verified. I also let the interviewees do the talking. According to Bogdan & Biklen (1998) a good interview is one that belongs mostly to the interviewee. The transcripts can verify this.

3.4.5 Data collection timetable

Quantitative data collection spanned most of 2013 with the qualitative interviews taking place between January and April, 2014. The data collection timetable is outlined in Table 3.3.

Table 3.3
Data collection timetable

2013	
February	Attended NZCETA meetings to inform and generate interest about the project
26 March	Ethics approved
June	Approached NZCETA head office to assist in contacting economics and accounting teachers.
4th July	NZCETA email newsletter to members with information regarding the research and a link to the survey.
19th July	19 responses to the survey
9th August	Ethics approved to offer incentive to participate in the survey
15th October	57 responses
29th October	All New Zealand secondary schools emailed directly and asked for the email to be forwarded to the relevant teacher
29th October	Link to the survey sent out on Google list serves for economics and accounting teachers by list serve owners
8th November	135 useable surveys
2014	
27th January – 10th April	All interviews were conducted

3.5 Ethical considerations

Ethics approval was obtained from the Victoria University of Wellington Faculty of Education Ethics Committee on 26th March, 2013: ethics approval no. SEPI/2013/10 RM RM 19669 (Appendices 3 & 4). All participants were recruited by invitation, and participation was voluntary. Survey participation was anonymous, unless the participant willingly shared contact details. The interviewees were not anonymous but were confidential, and informed consent was obtained. All participants were allocated a pseudonym to preserve confidentiality, and their schools have not been identified in this thesis. All opinions and data have been reported in an aggregated form and individuals are not identifiable. The data were stored on my password-protected computer, and transcribers signed a confidentiality agreement.

3.6 Quantitative data analysis

3.6.1. Factor analysis

Factor analysis is a statistical procedure for reducing the dimensionality of data and to investigate any apparent underlying constructs that manifest statistically as *factors* (Walkey & Welch, 2010). The key concept of factor analysis is that observed variables associated with the same latent variable (factor) will be correlated with one another more highly than they are with other observed variables. If items are associated sufficiently with the same factor they may be presumed to be (valid) measures of the same latent variable. Therefore, factor analysis can be used to aid understanding of the structure of correlations among items (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Ultimately, factor analysis helps to develop and refine models and theories of behaviour (Walkey & Welch, 2010).

Confirmatory factor analysis

Confirmatory factor analysis (CFA) is a method of assessing the adequacy of a theoretical model to explain a set of empirical data (Walkey & Welch, 2010). As the data for this study were gathered using a questionnaire based on an *a priori* model (G.T.L. Brown, 2006), CFA was carried out using AMOS software to evaluate the fit of the data to this model.

A variety of fit indices have been developed to determine how well a set of data fits a model. Kline (2010) recommended reporting the chi-squared test, the RMSEA (Root Mean Square of Error Approximation), the CFI (Comparative Fit Index), and the SRMR (Standardised Root Mean Square Residual).

Chi square test

Chi-square is used for evaluating overall model fit (Hooper, Coughlan, & Mullen, 2008) and “assesses the magnitude of discrepancy between the sample and fitted covariances matrices” (Hu & Bentler, 1999, p. 2). However, chi-square is sensitive to sample size and nearly always rejects models when sample sizes are large (Hooper et al.). Conversely, researchers may accept an inappropriate model if the sample size is small (Gatignon, 2014) and with only 135 respondents the chi-square test may not be reliable for the present data. To be acceptable, the chi-square value should indicate a non-significant difference between the observed data and those expected for the model. Tabachnick & Fidell (2007) recommended the use of chi-square per degrees of freedom (χ^2/df) to counteract the effect of sample size, and if the value is less than two, the data and the model are considered similar and the model a good fit (Walkey & Welch, 2010).

Root Mean Square of Error Approximation

To remove the population sample effect, RMSEA can be used (Steiger & Lind, 1980, as cited by Steiger, 1990). The RMSEA is an index of the discrepancy between the model and the data per degree of freedom for the model. The RMSEA is regarded as an informative fit index because of its sensitivity to the number of estimated parameters in the model (Hooper et al., 2008). According to Hooper et al., the cut-off points for RSMEA have fallen from a range of 0.05 to 0.10 being considered a fair fit to Steiger’s (2007) recommendation of a stringent upper limit of 0.07. A smaller RMSEA indicates a better fit of the model to the data, and a RMSEA equal to zero would indicate that the target model fits the data perfectly.

Comparative Fit Index

The CFI was introduced by Bentler (1990) and also takes sample size into account. It performs well with small samples (Hooper et al., 2008). The CFI compares the fit of a target model to the fit of an independent model (Bentler, 1990). Fan, Thompson, and Wang (1999) claimed it was the fit index least affected by sample size and again, opinion has varied on optimum cut-off, with Hu and Bentler (1999) regarding 0.95 as indicative of a good fit. A larger value of CFI indicates a better fit to the model.

Standardised Root Mean Square Residual

The SRMR is the square root of the difference between the residuals of the sample correlation matrix and the hypothesised correlation (Hooper et al., 2008). The discrepancy between the sample correlation and the model correlation indicates the degree of fit between the model and the data. Generally, an SRMR of less than 0.8 indicates a good fit of the model (Hu & Bentler, 1999).

Cut-off values

Optimal cut-off values are debated in the literature with Hu & Bentler's (1999) suggestions being called into question by Sivo, Fan, Witta, and Willse (2006). Sivo et al. concluded optimal cut-off values vary with sample size, for example, for RSMEA the optimal value for a small sample of 150 is 0.06, but may decrease to 0.02 if sample size is greater than 2,500. Measurement indices are also reported in terms of "goodness of fit" and "badness of fit" with "goodness of fit" indices requiring high values preferably between 0.95 and 1.00, whereas "badness of fit" indices require small values, generally between 0.00 and 0.10. G.T.L. Brown (2011b) proposed a range of acceptable fit statistics, which is reproduced in Table 3.4, although as warned by Sivo et al., small sample sizes need to err on the side of caution in terms of deciding on model fit.

Table 3.4

Range of Acceptable Fit Indices In CFA

Decision	Goodness of Fit		Badness of Fit	
	p of χ^2/df	CFI	RMSEA	SRMR
Good	>.05	>.95	<.05	<.06
Acceptable	>.05	>.90	<.08	<.08
Marginal	>.01	.85-.89	<.10	
Poor	<.01	<.85	>.10	>.08

Reproduced from G.T.L. Brown (2011b)

Sivo et al. (2006) recommended with a sample size of 150, models with statistically non-significant χ^2 per *df*, an optimal value of around .06 for root mean square errors of approximation (RMSEA), 0.12 for standardised root mean residuals (SRMR) and .95 for the comparative fit index (CFI) be considered the optimal cut-off values.

Exploratory factor analysis

As will be discussed fully in Chapter 4, the confirmatory factor analysis did not meet the recommended criteria to conclude that the data fit G.T.L. Brown's (2006) model. Brown and Remesal (2012) suggested that if models are inadmissible or poorly fitting, the model can be modified or exploratory factor analysis (EFA) carried out to develop a new model. Fabrigar et al. (1999) claimed there are five methodological issues to be considered when carrying out EFA:

1. Which variables to include in the study and what is the nature of the sample?
2. Is EFA the most appropriate form of analysis?
3. Which fitting procedure will be used?
4. Which method of rotation will be chosen?
5. How many factors should be extracted?

Variables included and nature of the sample

There are some guidelines with regard to adequate sample size. Gorsuch (1983) recommended at least five participants per measured variable and not less than 100 overall while Comfrey and Lee (1992) suggested a sample of 300 was good, 1,000 as excellent and 100 as poor. However, there are limitations to sample size guidelines, Fabrigar et al. (1999) claimed sample size is not a function of the measured variables *per se* and that overly homogenous samples should be avoided. Thus, by these criteria, EFA is appropriate. At $n = 135$, the present sample size is just adequate by Gorsuch's criterion; there were exactly five participants per survey item and more than the absolute minimum of 100 participants.

Conditions for EFA

The Kaiser-Meyer-Olkin (KMO) measure of sample adequacy can be calculated and values closest to one are desirable. Hutcheson (1999) suggested that a KMO over 0.60 is acceptable. A further measure to determine if there are correlations between the variables is Bartlett's test of sphericity, which indicates "whether the correlation matrix is significantly different from an identity matrix" (Field, 2013, p. 685). If the *Sig.* value for this analysis is less than 0.05 then the null hypothesis can be rejected and we can conclude there are correlations in the data that are appropriate for factor analysis.

The data in this study have a Kaiser-Meyer-Olkin (KMO) of 0.823 and Bartlett's test of sphericity (*Sig* < 0.001), and so EFA was considered an appropriate type of analysis.

Fitting procedure

A number of model-fitting methods are available in EFA, but two of the most commonly used approaches are Maximum Likelihood and Principal Axis Factoring (Yong & Pearce, 2013). Osborne, Costello, and Kellow (2008) recommended either of these two methods would give the optimal results. Maximum likelihood estimation was used for the data in this thesis.

Method of rotation

Factors are often rotated to remove ambiguity and to simplify the structure. Rotation can be orthogonal or oblique (Yong & Pearce, 2013). Orthogonal rotation constrains the factors to being uncorrelated, while oblique rotation allows for the factors to be correlated and for the axes to rotate freely “rather than being fixed at right angles to each other” (Walkey & Welch, 2010, p. 81). Osborne et al. (2008) recommended using an oblique rotation in social research because we can expect some correlation among factors, and according to Walkey and Welch, oblique rotation may provide a better solution and “a better representation of the factor structure” (p. 81). Field (2013) also recommended oblique rotation when theory suggests that the constructs relating to the factors might correlate and Osborne et al. suggested an orthogonal rotation might result in lost information if the factors are really correlated and suggested an oblique rotation should “render a more accurate and perhaps more reproducible solution” (p. 90).

An oblique rotation was chosen for the data in this research because G.T.L. Brown’s (2006) model had already established correlations among the variables and factors. Within oblique rotation there are three methods of rotation (direct oblimin, quartimin and promax), although according to Fabrigar et al. (1999), all three tend to produce similar results. Yong and Pearce (2013) suggested that Promax may result in greater correlations, therefore Promax was chosen as the method of rotation for the data.

Extracting the factors

Researchers must decide how many factors to extract as not all possible factors are typically retained (Field, 2013), and in choosing the number of factors they must balance parsimony with plausibility (Fabrigar et al., 1999). There are a number of procedures that can be used when deciding on the number of factors to retain. Three commonly used are the Kaiser criterion of retaining all factors with an eigen value greater than one, the scree test where the scree plot is examined to determine the point of inflexion and where the slope of the line changes the number of factors above this point are retained (Field, 2013), and parallel analysis which is a method based on the generation of random variables, to determine the number of factors to retain (Ledesma & Velero-Mora, 2007). According to Zwick and Velicer (1986), parallel analysis is

the most appropriate method for deciding the number of factors to retain, so this was carried out using O'Connor's (2000) syntax.

Summary of EFA

An initial EFA using Maximum Likelihood Estimation with an Oblique Rotation (Promax) using parallel analysis and the Scree plot was carried out on the 27 items. This was followed up with further EFA reducing and fixing the number of factors, and subsequent analyses involved removing items to establish an acceptable model.

3.7 Qualitative data analysis

Qualitative data analysis is the systematic arrangement of interview transcripts and field notes; it involves organising and synthesising data to search for patterns (Bogdan & Biklen, 1998). Sorting the data into patterns involves developing a coding system which includes clustering like items together. Codes are labels or categories to which units of data are assigned and units can consist of single words, phrases, sentences or whole paragraphs (Miles & Huberman, 1994). Codes can be either deductive, those previously developed, or inductive where codes are generated by the data (Teddlie & Tashakkori, 2009). Inductive coding requires the use of a constant comparative method (Glaser & Straus, 1967) which requires the comparison of data to be assigned a code with those already in the code, distinguishing differences between codes and integration of codes (Zhang & Wildemuth, 2009).

Thematic analysis of qualitative data involves examining the codes for broader patterns of meaning and collating data to each potential theme (Braun & Clarke, 2006). Themes need to be checked against the data set to ensure they answer the research questions, and then named and defined (Braun & Clarke).

Miles and Huberman (1994) suggested the following framework for analysing qualitative data:

- data reduction: simplifying and transforming the data into transcriptions which includes summarising and paraphrasing;
- data display: organising and compressing the data into matrices, charts, networks;
- drawing and verifying conclusions: deciding what the data means, noting patterns, explanations, propositions.

3.7.1 Data Reduction

The audio files of the interviews were transcribed into Word documents before being uploaded into NVivo 10. The initial coding process involved identifying ideas (made up of words, phrases and sentences) from interview transcripts and grouping like ideas together. These groupings of like ideas were paraphrased with codes and rechecked several times to ensure agreement with the initial codes, and that there was consistency when asked “what is this piece of data an example of”? Initial codes were analysed for underlying meaning and after several iterations five themes emerged to which data could be collated.

3.7.2 Data display

Data displays help to organise, compress and assemble information (Punch, 2005), therefore matrices and diagrams were used. Five themes were developed from the data and summarised in a diagrammatical overview.

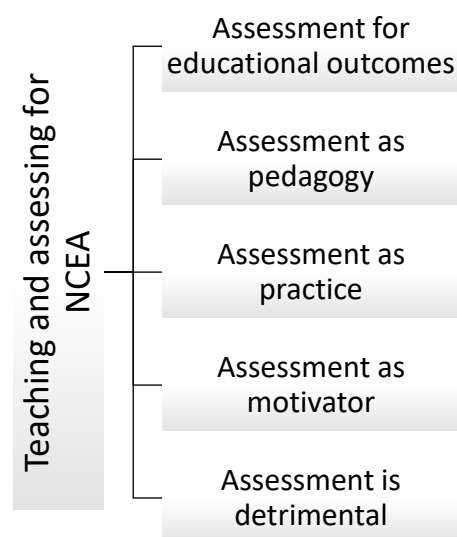


Figure 3.1. Overview of themes derived from qualitative data

Each theme was then explained and compressed in a matrix. Figure 3.2 illustrates part of the matrix for the theme “Assessment as pedagogy”. The left-hand side of the matrix is the theme with a description of the theme, while the right-hand side describes the data units assigned to that theme.

<p>Assessment as pedagogy</p> <p>This theme includes all comments that relate to assessment being, or is, part of the teaching and learning process.</p>	<ul style="list-style-type: none"> • Assessment informs student learning • Assessment informs my teaching • Internal assessment can be integrated with teaching, but this can be problematic • Distinguish between assessment in the junior and senior school
-----------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Figure 3.2. Example from matrix of themes and some codes contributing to that theme

3.7.3 Drawing and verifying conclusions

The development of five themes was an iterative process, carried out in conjunction with supervisors and tested with colleagues. Initial coding resulted in eight overarching codes with 21 sub-codes. For example, two initial codes were Course Design and Internal Assessment, but subsequent analysis of the data began to suggest that many teacher actions with regard to assessment and courses involved strategising to ensure student success and preparation for their next step – either educational or workforce. Therefore, the theme “Assessment for educational outcomes” was developed from these ideas.

Researchers can also develop typologies to help make sense of the data (Johnson & Christensen, 2014). Analysis of the qualitative data suggested participants’ beliefs and actions towards internal assessment for NCEA could be conceptualised along a continuum of maximising NCEA internal assessments for the benefit of students’ achievement to treating internal assessment as an extension of the external examinations. There seemed to be three dimensions to these actions and beliefs – the conditions under which internal assessment was conducted, the types of assessment tasks teachers asked students to complete, and an overall attitude towards internal assessment. These typologies are summarised in a matrix.

Teachers' beliefs about conducting internal assessment for NCEA

Teacher
(Ushers)

Examiner
(Gatekeepers)

Assessment conditions	Attitudes	Assessment tasks
<ul style="list-style-type: none"> • Open book • For mative feedback provided • Aut henticity assured through ongoing monitoring • Resub mssions 	<ul style="list-style-type: none"> • Positi ve towards NCEA and internal assess ment . • Internal assess ment should gi ve students more opportunities to achieve 	<ul style="list-style-type: none"> • Ori gi nal or large adapt ations to exemplars • Adapt ations to meet student needs • Portfolios • Cont ext ualised
<ul style="list-style-type: none"> • Exa mli ke conditions • No or mi ni mal feedback • No resub mssions • Aut henticity a dri ver and bi g concern 	<ul style="list-style-type: none"> • Negati ve towards NCEA and internal assess ment • Internal assess ment too easy • Ot her schools too generous 	<ul style="list-style-type: none"> • Mi nor adapt ations to exemplars • Adapti ons to ensure aut henticity

Figure 3.3. Typology of teachers' beliefs and actions towards NCEA internal assessment

3.8 Validity and reliability in a mixed methods design

3.8.1 Validity and reliability in quantitative data

In quantitative research, the concepts of validity and reliability are used as quality measures. Validity refers to the extent to which an instrument measures what it is intended to measure (Muijs, 2011). Abstract concepts such as conceptions of assessment cannot be directly measured, so an instrument to gather data *associated with* the concept is needed. A validated instrument, G.T.L. Brown's (2006) TCoAIIIA Inventory, was chosen to measure conceptions of assessment.

The reliability of an instrument refers to the consistency or stability of data measured using that instrument (Johnson & Christensen, 2014). Reliability can be calculated using a coefficient, such as Cronbach's alpha, which is an estimate of the overall reliability of a questionnaire. Cronbach's alpha estimates the degree to which items

are interrelated. A general 'rule of thumb' for reliability is that the coefficient alpha should be greater than 0.7 (Johnson & Christensen). Cronbach's alpha was calculated for each factor identified in the exploratory factor analysis.

3.8.2 Validity and reliability in qualitative data

Lincoln and Guba (1985) suggested that for qualitative data trustworthiness replaces validity and suggested four criteria to ensure this: credibility, transferability, dependability and conformability. They suggested a number of strategies to ensure credibility, which include prolonged engagement, persistent observation, triangulation, peer debriefing, negative case analysis, referential adequacy, and member checking. In this study, the strategies of prolonged engagement, triangulation, peer debriefing, negative case analysis, and member-checking were used.

Prolonged engagement can refer to speaking with a range of people (Lincoln & Guba, 1985) and was achieved through interviewing 21 participants from a wide variety of settings. Triangulation is the use of more than one data source. In this research, interview participants also completed the questionnaire, enabling comparisons between the two data sources. Peer debriefing is the process of discussing findings with peers to uncover biases, assumptions, and perspectives (Lincoln & Guba) and this was carried out with my supervisors and colleagues who were formerly teachers in the NCEA assessment regime. Negative case analysis is discussing elements of the data that do not support general patterns and again this was carried out with my supervisors. Member checking involves inviting participants to comment on transcriptions and interpretations, to ensure that these were conveyed accurately, so all transcripts were returned to participants for checking and all agreed with their original comments.

Lincoln and Guba (1985) define transferability as showing that the findings have applicability elsewhere and claimed that this can be gained by describing the phenomena in sufficient detail for readers to determine whether the findings are transferable to other settings. Dependability determines whether the findings and conclusions are supported by the data and involves having another researcher examine the process and the product of the study which in this case was done by supervisors.

Confirmability refers to the extent to which the findings are shaped by the participants and not researcher bias. Lincoln and Guba claim this can be achieved through audit, triangulation, or reflexivity. This study was audited by my supervisors and triangulated through mixed methods.

3.8.3 The inference process

Quantitative and qualitative approaches have their own criteria for evaluation, and Creswell (2012) stated that researchers must be mindful of these at all stages of the study. However, Heyvaert, Hannes, Maes, and Onghena (2013) argue that merely attending to the criteria for each strand is not enough “because an MM study is more than simply the sum of its qualitative and quantitative elements” (p. 316). Therefore, attention must be paid to the overarching analysis as the strands in mixed methods research do not just answer strand-specific questions; each should contribute to overall research questions.

According to Teddlie and Tashakkori (2009), “the main reason for using a mixed methods approach is to provide better understanding of the phenomenon under investigation” (p. 286). However, as they say, enhanced understanding is only possible if the outcomes, that is the conclusions and interpretations, of both strands (quantitative and qualitative) are effectively integrated. Teddlie and Tashakkori refer to these outcomes of research as inferences where inferences denote a process and an outcome. They describe three related concepts:

- the inference process – making sense of the results of data analysis;
- inference quality – the standards for evaluating the quality of conclusions. In quantitative terms internal validity and in qualitative terms credibility and trustworthiness;
- inference transferability – the degree to which conclusions may be applied to other settings (generalisability and transferability).

Teddle and Tashakkori (2009) provided some guidelines to make inferences that are credible and comprehensive. These include getting to know the participants, having a deep understanding of the literature, analysing data with the explicit intention of answering research questions, using respective quality criteria for the quantitative and qualitative strands, and integrating inferences from each strand into a credible meta-inference.

In 2010, O’Cathain, Murphy, and Nicoll provided a framework to assess the quality of mixed methods approaches in medical research, but their techniques also have relevance to educational research. The three techniques they suggested are using a triangulation protocol, following a thread, and a mixed methods matrix. To carry out the process of triangulation they suggest creating a matrix of findings from each part of the study and comparing for convergence, complementary information, and discrepancies. Following a thread refers to researchers selecting a theme from one component and following it across the other; however, this method tends to be more relevant when interviews are used to generate a hypothesis which is then tested in surveys. They also suggested that a mixed methods matrix can be used to compare data from both sources for the same participants, focusing on cases rather than themes.

Ivankova (2014) added to the discussion on quality assurance in mixed methods by specifically dealing with a sequential QUAN → QUAL design and provided a three-step process for securing quality of the meta-inferences. These are: a systematic process for selecting participants for the QUAL part of the study; elaborating on unexpected QUAN results; and observing the interaction between the two strands of the study. Creswell (2012) informed researchers using mixed methods that they must interpret their combined results, and for an explanatory sequential design that a display linking qualitative themes with quantitative results can be used.

To observe interaction between the qualitative and quantitative strands a matrix to display convergence and complements was constructed.

Quantitative factors	Overarching conceptions	Qualitative themes
Assessment for improvement	Assessment is for learning Assessment for improvement Assessment as pedagogy Assessment as motivator	Assessment as pedagogy Assessment as a motivator
Assessment describes learning	Assessment is for qualifications Assessment for educational outcomes Assessment describes learning	Assessment for educational outcomes
Assessment has weaknesses	Assessment is detrimental Assessment has weaknesses Assessment as detrimental	Assessment is detrimental
Assessment for school accountability	Assessment is for accountability Assessment as practice Assessment for school accountability	Assessment as practice

Figure 3.4. Convergence and complements among quantitative factors and qualitative themes

CHAPTER 4

Quantitative Findings

“All models are wrong, but some are useful” (George Box, 1987, p. 424).

4.1 Confirmatory analysis

Confirmatory factor analysis was carried out to evaluate the fit of the data to G.T.L. Brown’s (2006) model, but the analyses showed an inadmissible fit. Table 4.1 shows the fit indices for the data in this research. While the chi-squared-per-degrees-of-freedom and RMSEA statistics show acceptable fit, the other fit statistics suggest rejecting the model. Chi-square per degrees of freedom was 1.729 and the RMSEA was 0.074. However, with the CFI at .83 and SRMR at .0848 the data collected for this research are considered a marginal-to-poor fit to G.T.L. Brown’s model.

Table 4.1
Fit Indices for 135 Secondary Teachers of Accounting and Economics

Model		Fit statistics			
<i>n</i>	Group	χ^2/df	CFI	RMSEA	SRMR
135	Secondary teachers of accounting and economics	537.674/311 = 1.729 <i>p</i> of χ^2/df = .19	0.835	0.074	.0848

In addition to examining fit statistics, Bowen and Guo (2011) recommend that researchers should also examine factor loadings for strength of loading. They did note that acceptable loadings are not as clearly defined as those for EFA, but suggested rejecting loadings below one of the common EFA cut-offs, for example those below 0.4, a figure with which Bandalos and Finney (2010) agreed. Figure 4.1 illustrates factor loadings from the CFA.

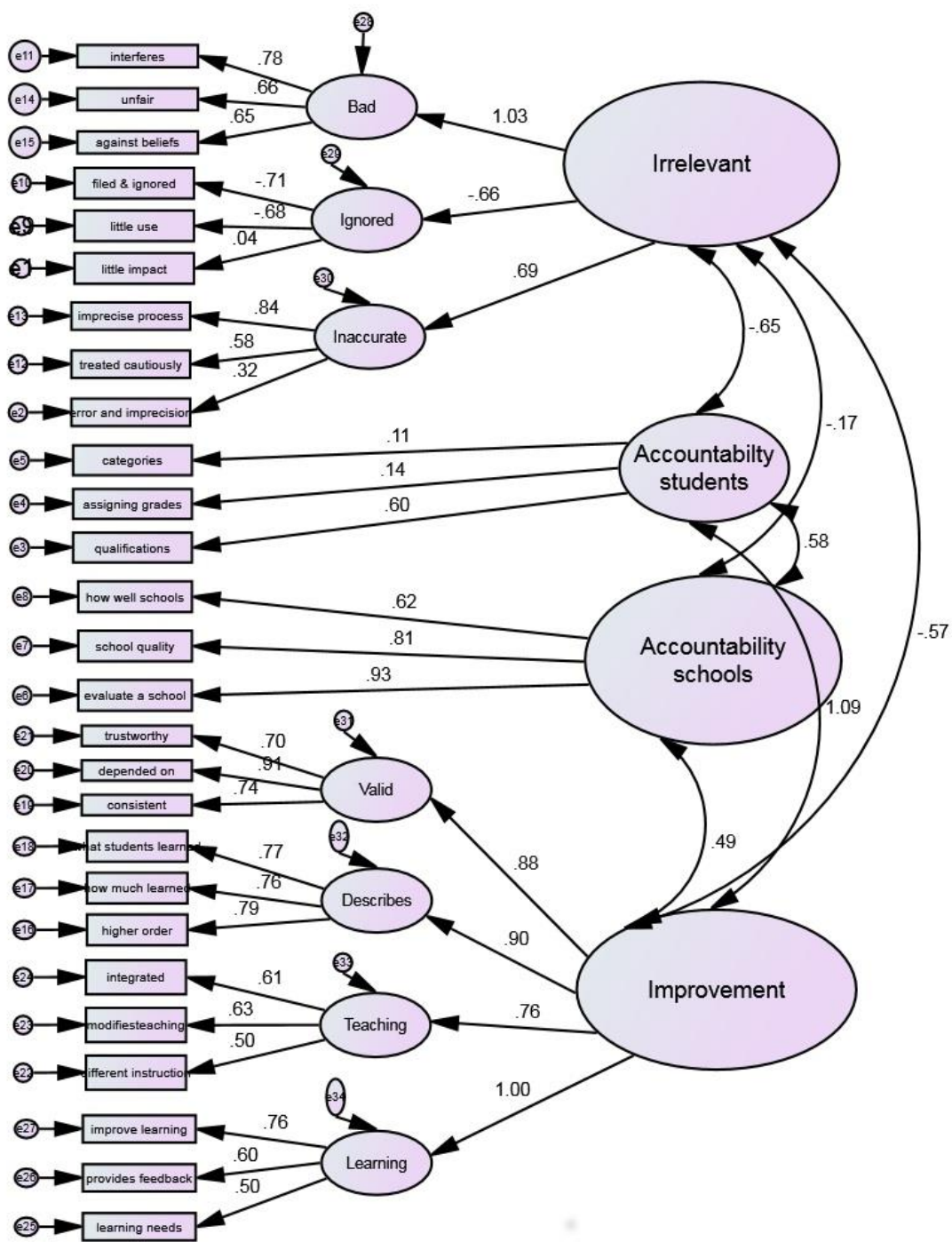


Figure 4.1. Measurement model of 135 economics and accounting teachers' conceptions of assessment

4.1.1 Strength of the present data loadings to factors

While many loadings are above the acceptable weighting, some are below and others are greater than one. According to T.A. Brown (2006), the presence of such out-of-range values, or Heywood cases, indicates that the model is not acceptable. However, Joreskog (1999) argued that with an oblique rotation the factor loadings are regression coefficients and not correlations “and as such can be greater than one in magnitude” (p. 1) and do not necessarily mean something is wrong. Bentler and Chou (1987) pointed out that such estimation errors are more likely with small samples.

Assessment improves education second order factor

Table 4.2 shows the first order factors that in G.T.L. Brown’s (2006) model were associated with the *assessment improves education* second order factor.

Table 4.2
Items relating to first order factors which contribute to second order factor of “assessment improves education”

First order factors		Items loading to the factor	Strength of loading to factor
Assessment is valid	Question 6	Assessment results are trustworthy	.70
	Question 24	Assessment results can be depended on	.91
	Question 15	Assessment results are consistent	.74
Assessment describes abilities	Question 12	Assessment establishes what students have learned	.77
	Question 3	Assessment is a way to determine how much students have learned from teaching	.76
	Question 21	Assessment measures students’ higher order thinking skills	.79
Assessment improves teaching	Question 5	Assessment is integrated with teaching practice	.61
	Question 14	Assessment information modifies ongoing teaching of students	.63
	Question 23	Assessment allows different students to get different instruction	.50
Assessment improves learning	Question 22	Assessment helps students improve their learning	.76
	Question 4	Assessment provides feedback to students about their performance	.60
	Question 13	Assessment feeds back to students their learning needs	.50

All items are moderately to strongly associated with their first order factors. There are also strong correlations between most of the first and second order factors. However, the correlation between the first order factor *assessment improves learning* and the second order factor *assessment improves education* is problematic at 1.0 (see Figure 4.1); Fabrigar et al. (1999) warned that this indicates issues with the data and the model fit.

Assessment makes schools accountable first order factor

In G.T.L. Brown’s (2006) model, three items contribute to the first order factor of *assessment makes schools accountable*, shown on Table 4.3.

Table 4.3
Items relating to first order factor of “Assessment Makes Schools Accountable”

First order factor	Items loading to the factor		Strength of loading to factor
Assessment makes schools accountable	Question 1	Assessment provides information on how well schools are doing	.62
	Question 10	Assessment is an accurate indicator of a school’s quality	.81
	Question 19	Assessment is a good way to evaluate a school	.93

The present data show that the items load moderately to strongly to this factor at .62, .81 and .93 respectively.

Assessment makes students accountable first order factor

According to G.T.L. Brown’s (2006) model, three items contribute to the first order factor of *assessment makes students accountable*, shown on Table 4.4.

Table 4.4
Items relating to first order factor of “Assessment Makes Students Accountable”

First order factor	Items loading to the factor		Strength of loading to factor
Assessment makes students accountable	Question 2	Assessment places students in categories	.11
	Question 11	Assessment is assigning a grade or level to students	.14
	Question 20	Assessment determines if students meet qualifications standards	.60

The data show items 2 and 11 are loading weakly to the student accountability factor at .11 and .14 (respectively), while item 20 loads strongly at .60.

Assessment is irrelevant second order factor

G.T.L. Brown’s (2006) second order factor of *assessment is irrelevant* consists of three first order factors: *assessment is bad*; *assessment is ignored*; *assessment is inaccurate*. The items contributing to each of the first order factors are in Table 4.5.

Table 4.5
Items relating to first order factors which contribute to second order factor of “Assessment is Irrelevant”

First order factors	Items loading to the factor		Strength of loading to factor
Assessment is bad	Question 25	Assessment interferes with teaching	.78
	Question 16	Assessment is unfair to students	.66
	Question 7	Assessment forces teachers to teach in a way that is against their beliefs	.65
Assessment is ignored	Question 17	Assessment results are filed & ignored	.71
	Question 8	Teachers conduct assessments but make little use of the results	-.68
	Question 26	Assessment has little impact on teaching	.04
Assessment is inaccurate	Question 27	Assessment is an imprecise process	.84
	Question 9	Assessment results should be treated cautiously because of measurement error	.58
	Question 18	Teachers should take into account the error and imprecision in all assessment	.32

The data show that the first three items relating to the *assessment is bad* factor load moderately to that factor, at .78, .66 and .65. Item 26, *assessment had little impact on teaching*, loaded weakly to the *assessment is ignored* factor. One item (27) loaded well to the *assessment is inaccurate* factor at .84, but item 9 had a moderate loading at .58, while item 18 loaded below the recommended level at .32.

Loading between first and second order factors

The first order factor of *assessment is ignored* correlated inversely with the second order factor *assessment is irrelevant*, whereas in G. T. L. Brown (2006) this was a positive correlation. Also the out-of-range (Heywood) loading between the first order factor of *assessment is bad* and the second order factor of *assessment is irrelevant* is problematic; it might indicate a mis-specified model or that “the data violate the assumptions of the common factor model” (Fabrigar et al., 1999, p. 276).

4.2 Acceptability of G.T.L. Brown’s (2006) model

Overall, the fit statistics indicate that the original model of G.T.L. Brown (2006) should be rejected, but on closer examination of the CFA model (Figure 4.1) some aspects of the original model have been reproduced. Furthermore, inter-correlations among the factors have similarities with G.T.L. Brown’s model that showed: the student and school accountability factors were positively correlated; the improvement factor was positively correlated with school and student accountability factors; and the irrelevant factor inversely correlated to the improvement factor but had a statistically insignificant correlation with the school accountability factor.

G.T.L. Brown (2011a) reported a study that included New Zealand secondary teachers and which showed similar inter-correlations with his previous work undertaken solely with primary teachers. In both G.T.L. Brown’s (2011a) research and the present study, the improvement factor correlated positively with school and student accountability, but inversely with the irrelevant conception. School and student accountability were positively correlated, and irrelevance and school accountability inversely correlated. It should be noted, however, for the present data, the correlation between improvement and student accountability is greater than one and therefore problematic.

An important difference from G.T.L. Brown’s (2011a) research is the inverse correlation between the irrelevant factor and the student accountability factor. G.T.L. Brown (2011a) explained the positive correlation between these two factors as showing that “the more teachers agreed that assessment was irrelevant, the more they saw it as a means of grading students summatively” (p. 113). He argued that for primary teachers, making students accountable was linked to irrelevance. However, for secondary teachers assessing students for summative accountability is not an irrelevant use of assessment because it has consequences for students’ attainment of qualifications, a contextual factor of secondary but not primary education. This may explain the inverse correlation in the present data. Table 4.6, conceptions of assessment inter-correlations: New Zealand primary and secondary teachers (G.T.L. Brown, 2011a) compared with conceptions of assessment inter-correlations: economics and accounting teachers, shows the similarities and differences to G.T.L. Brown’s model.

Table 4.6

Conceptions of assessment inter-correlations: New Zealand primary and secondary teachers (G.T.L. Brown, 2011a) compared with economics and accounting teachers

Concepti ons	G.T.L. Brown’s model				Present data			
	1	2	3	4	1	2	3	4
1. Impr ove nent	-	-.74	.47	.23	-	-.57	.49	1.09
2. Irreleva nce		-	-.14	.30		-	-.17	-.65
3. School Accountabi lity			-	.45			-	.58
4. Student Accountabi lity				-				-

The relatively small sample size of 135 might also be part of the explanation for the data not fitting models previously reported by G.T.L. Brown (2006, 2011a). However, the group of participants in this study is distinctive in that their main, sometimes sole, focus is teaching specific subjects in the senior secondary school and awarding NCEA qualifications. Therefore, it is also possible that their conceptions of assessment differ from participants in previous studies.

4.3 Exploratory factor analysis

Because of the poor fit of the data with G.T.L. Brown's (2006) model, EFA was carried out to attempt to find a better fitting model. An initial analysis using Maximum Likelihood Estimation with an Oblique Rotation (Promax) was carried out on the 27 items. An oblique rotation was used because this type of rotation allows for correlation among the factors and, as Osborne et al. (2008) stated, we can expect correlation among factors in the social sciences. The KMO index at 0.823 and Bartlett's test of sphericity ($Sig = 0.000$) verified that the sample was adequate for EFA. The initial analysis identified eight factors with Eigen-values greater than one, whereas the point of inflexion on the Scree plot suggested four factors be retained, as did the parallel analysis using O'Connor's (2000) syntax.

4.3.1 Four factor analysis

The Factor Matrix (Appendix 5) for a four factor solution showed four items not loading to any factor (Item 2 – assessment places students in categories, Item 11 – assessment is assigning a grade, Item 18 – teachers should take into account error and imprecision in all assessment, and Item 26 – assessment has little impact on teaching). The anti-image correlation matrix (Appendix 6) showed two variables (Items 18 & 26) having values less than 0.5. The diagonal of the anti-image correlation matrix shows the KMO values for each individual variable, and Field (2013) recommended that all values in the anti-image correlation matrix should be above a bare minimum of 0.5. The Pattern Matrix, Table 4.7, shows three items with cross-loadings, but all four factors have at least three items.

Table 4.7

Pattern matrix extracted using Maximum Likelihood Estimation with an Oblique rotation (Promax) with a four-factor solution

	Factor			
	1	2	3	4
Assessment is a way to determine how much students have learned from teaching	.841			
Assessment determines if students meet qualifications standards	.763			
Assessment results can be depended on	.753			
Assessment helps students improve their learning	.710			
Assessment establishes what students have learned	.709			
Assessment results are trustworthy	.686			
Assessment measures students' higher order thinking skills	.665			
Assessment results are consistent	.581			
Assessment is unfair to students	-.534	.321		
Assessment results are filed & ignored	-.449			
Assessment allows different students to get different instruction	.433			
Assessment is integrated with teaching practice	.327			.323
Assessment is an imprecise process		.743		
Assessment forces teachers to teach in a way against their beliefs		.730		
Assessment results should be treated cautiously because of measurement error		.636		
Assessment interferes with teaching		.627		
Teachers conduct assessments but make little use of the results		.486		
Assessment is a good way to evaluate a school			.897	
Assessment is an accurate indicator of a school's quality			.823	
Assessment provides information on how well schools are doing			.478	
Assessment places students into categories				
Assessment information modifies ongoing teaching of students				.769
Assessment provides feedback to students about their performance	.417			.477
Assessment feeds back to students their learning needs				.396
Teachers should take into account the error and imprecision in all assessment				
Assessment has little impact on teaching				
Assessment is assigning a grade or level to student work				
Extraction Method: Maximum Likelihood.				
Rotation Method: Promax with Kaiser Normalization.				
a. Rotation converged in 6 iterations.				
Eigen values	7.669	2.614	1.93	1.61
% of variance	28.4	9.68	7.17	5.98

Osborne et al. (2008) and Yong and Pearce (2013) recommend that multiple factor analyses be run at numbers above and below the predicted number of factors. So, analyses with three and five factors were also run to see if that produced a more appropriate result.

4.3.2 Three factor analysis

The Factor Matrix for a three factor solution showed the same four items not loading to any factors (Item 2 – assessment places students in categories, Item 11 – assessment is assigning a grade, Item 18 – teachers should take into account error and imprecision in all assessment, and Item 26 – assessment has little impact on teaching). The anti-image correlation matrix also showed two variables (Items 18 & 26) having values less than 0.5. The Pattern Matrix, Table 4.8, does show all factors with at least three items loading and only one item cross-loading.

Table 4.8

Pattern matrix extracted using Maximum Likelihood Estimation with an Oblique rotation (Promax) with a three-factor solution

	Factor		
	1	2	3
Assessment is a way to determine how much students have learned from teaching	.751		
Assessment helps students improve their learning	.735		
Assessment results can be depended on	.703		
Assessment provides feedback to students about their performance	.687		
Assessment measures students' higher order thinking skills	.640		
Assessment results are consistent	.630		
Assessment determines if students meet qualifications standards	.622		
Assessment allows different students to get different instruction	.603		
Assessment establishes what students have learned	.586		
Assessment results are trustworthy	.554		
Assessment is integrated with teaching practice	.518		
Assessment information modifies ongoing teaching of students	.496		
Assessment is unfair to students	-.449		.337
Assessment feeds back to students their learning needs	.443		
Assessment results are filed & ignored	-.399		
Assessment is a good way to evaluate a school		.933	
Assessment is an accurate indicator of a school's quality		.915	
Assessment provides information on how well schools are doing		.546	
Assessment places students into categories			
Assessment has little impact on teaching			
Assessment is assigning a grade or level to student work			
Assessment forces teachers to teach in a way against their beliefs			.723
Assessment is an imprecise process			.714
Assessment interferes with teaching			.630
Assessment results should be treated cautiously because of measurement error			.588
Teachers conduct assessments but make little use of the results			.459
Teachers should take into account the error and imprecision in all assessment			
Extraction Method: Maximum Likelihood.			
Rotation Method: Promax with Kaiser Normalization.			
a. Rotation converged in 6 iterations.			
Eigen values	7.669	2.614	1.93
% of variance	28.4	9.68	7.17

4.3.3 Five factor analysis

A five factor analysis was also carried out but the Factor Matrix also showed the same four items not loading to any factor and the anti-image correlation matrix with the same two variables (Items 18 & 26) having values less than 0.5. This analysis also produced six items with cross-loadings, and a number of items loading below the recommended .4 weighting, but factors had at least three items.

4.3.4 Removing items

If the results are still difficult to interpret, Costello and Osborne (2005) recommended removing problematic items; however, researchers have to consider whether this will compromise the data. G.T.L. sBrown (2011b) also warned that such modifications have to be theoretically defensible. In considering each extraction, the same four items did not load to any factors (Item 2 – assessment places students in categories, Item 11 – assessment is assigning a grade, Item 18 – teachers should take into account error and imprecision in all assessment, and Item 26 – assessment has little impact on teaching). As both the three-factor analysis and the four-factor analysis produced a similar number of issues, it was decided to follow the scree plot and parallel analysis recommendations of extracting four factors.

The analyses was run again with a four-factor extraction removing these items one by one. After multiple analyses, a four-factor solution with items 2, 11, 18 and 26 removed resulted in the most plausible structure. The pattern matrix, Table 4.9, shows all four factors have at least three items with loadings of an acceptable magnitude and the Cronbach's alpha for each factor is within an acceptable range as defined by George and Mallery (2003). There are, however, three items cross-loading to more than one factor.

Table 4.9

Pattern matrix extracted using Maximum Likelihood Estimation with an Oblique rotation (Promax) with a four-factor solution and items 2, 11, 18 & 26 removed

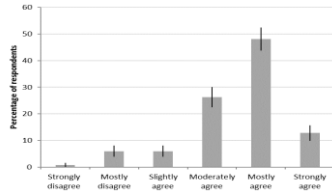
	Factor			
	1	2	3	4
Assessment determines if students meet qualifications standards	.787			
Assessment is a way to determine how much students have learned from teaching	.770			
Assessment results can be depended on	.722			
Assessment results are trustworthy	.700			
Assessment establishes what students have learned	.665			
Assessment helps students improve their learning	.645			
Assessment measures students' higher order thinking skills	.611			
Assessment is unfair to students	-.532	.335		
Assessment results are consistent	.531			
Assessment results are filed & ignored	-.413			
Assessment forces teachers to teach in a way against their beliefs		.727		
Assessment is an imprecise process		.707		
Assessment interferes with teaching		.642		
Assessment results should be treated cautiously because of measurement error		.585		
Teachers conduct assessments but make little use of the results		.475		
Assessment information modifies ongoing teaching of students			.881	
Assessment provides feedback to students about their performance			.568	
Assessment feeds back to students their learning needs			.482	
Assessment is integrated with teaching practice			.366	
Assessment allows different students to get different instruction	.310		.346	
Assessment is a good way to evaluate a school				.892
Assessment is an accurate indicator of a school's quality				.835
Assessment provides information on how well schools are doing	.305			.472
Extraction Method: Maximum Likelihood.				
Rotation Method: Promax with Kaiser Normalization.				
a. Rotation converged in 6 iterations.				
Eigen values	7.669	2.614	1.93	1.61
% of variance	28.4	9.68	7.17	5.98
Cronbach's alpha	.730	.766	.741	.813

4.3.5 Factors extracted

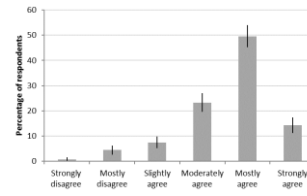
Each factor may be associated with a latent variable. The items loading to Factor 1 seem to describe actions relating to a summative function of assessment. The co-dimensionality of *assessment determines if students meet qualifications standards* with items such as *assessment measures students higher order thinking skills*, *establishes what they have learned*, and that assessment results are trustworthy, dependable and consistent, but assessment is not ignored, nor unfair, suggests a legitimate summative function of assessment. The association of *assessment helps students improve their learning* with these other items could indicate that participants believe summative functions of assessment can have a formative use. This could possibly relate to their role as assessors of NCEA where teachers award qualifications but are aware that this process may also help student learning.

The survey responses to the items contributing to this factor show general agreement with *assessment determines if students meet qualifications standards* and is a way to determine how much students have learned from teaching. There is less agreement that results are trustworthy, consistent and can be relied upon, and that assessment helps students improve their learning. Participants did not agree that assessment is unfair to students, nor that results are filed and ignored. This is a logical association because, if assessment has a useful summative function then it cannot be unfair or ignored. The responses to each item are in Figure 4.2.

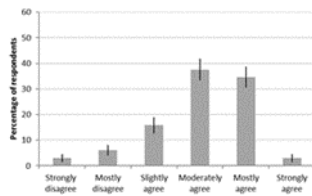
Assessment determines if students meet qualifications standard



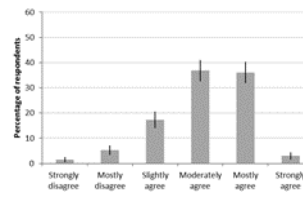
Assessment is a way to determine how much students have learned from teaching



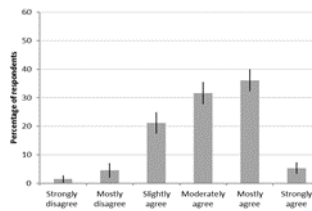
Assessment results are trustworthy



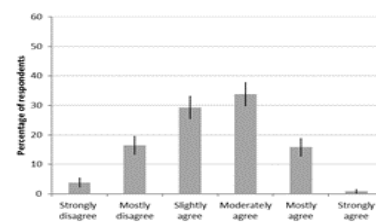
Assessment establishes what students have learned



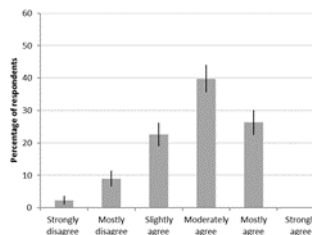
Assessment measures students' higher order thinking skills



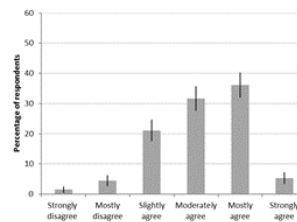
Assessment results are consistent



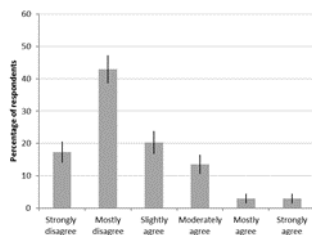
Assessment results can be depended on



Assessment helps students improve their learning



Assessment results are filed and ignored



Assessment is unfair to students

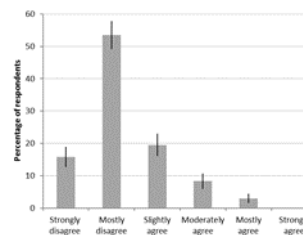
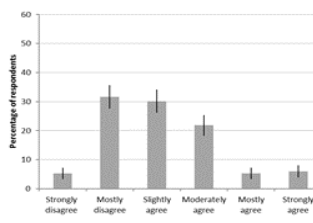
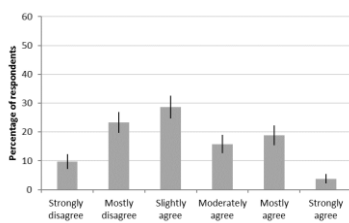


Figure 4.2. Distribution of responses from items contributing to factor – assessment describes learning.

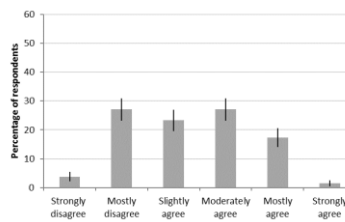
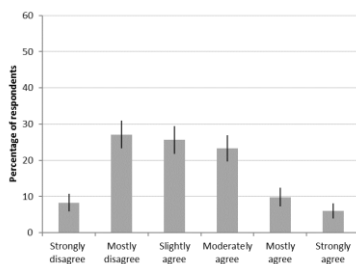
In G.T.L. Brown's (2006) model, the items loading on Factor 2 relate to the concept of assessment being detrimental to teaching and learning. However, participants were not in general agreement with these items. There is only some tendency towards believing assessment has detrimental aspects, therefore this factor has been named – Assessment has weaknesses.

Assessment forces teachers to teach in a way against their beliefs *Assessment is an imprecise process*



Assessment interferes with teaching

Assessment results should be treated cautiously because of measurement error



Teachers conduct assessments but make little use of the results

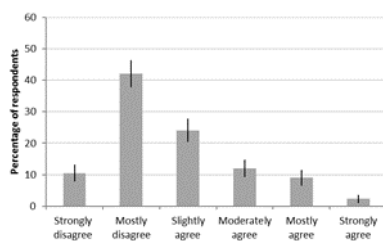
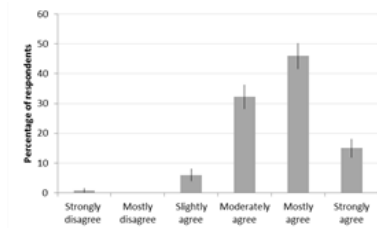


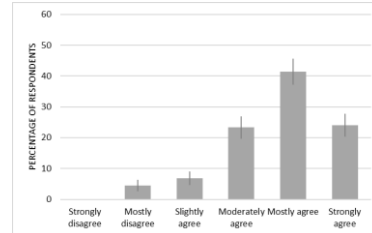
Figure 4.3. Distribution of responses from items contributing to factor – assessment has weaknesses

Items loading to Factor 3 all relate to the idea that assessment can improve learning and teaching and therefore has formative functions. The individual items show general agreement with formative functions of assessment, apart from item 23 – *assessment allows different students to get different instruction*.

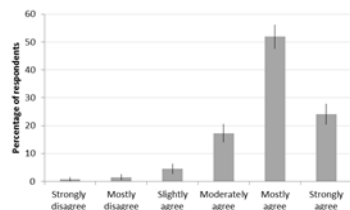
Assessment information modifies ongoing teaching of students



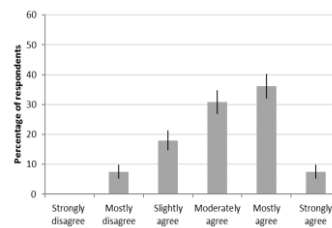
Assessment is integrated with teaching practice



Assessment provides feedback to students about their performance



Assessment feeds back to students their learning needs



Assessment allows different students to get different instruction

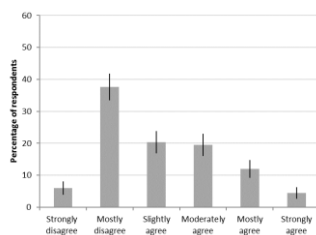


Figure 4.4. Distribution of responses from items contributing to factor – assessment improves learning and teaching.

The items loading to Factor 4 clearly align with G.T.L. Brown’s (2006) conception of school accountability. However, responses to the items in this factor are varied and not necessarily showing general agreement that assessment results can be a measure of school quality.

Assessment provides information on how schools are doing

Assessment is an accurate indicator of a school’s quality

Assessment is a good way to evaluate a school

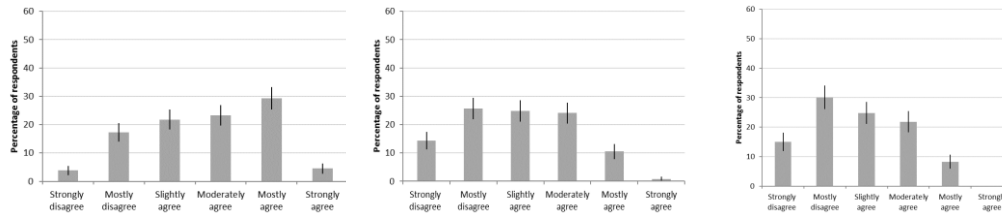


Figure 4.5. Distribution of responses from items contributing to factor – assessment for school accountability

The correlation matrix, Table 4.10, describes logical correlations. Factor 2 – *assessment has weaknesses* – is inversely correlated with all of the other factors. This is logical because if assessment has useful summative and formative functions and is possibly an indicator of school quality then its weaknesses should not be so great as to render the assessment unfit for its purposes. Factor 1 – *assessment describes learning* correlates positively with *assessment improves learning* and *assessment for school accountability*. This indicates that teachers agreed, somewhat, with the idea that assessment results could be an indicator of school quality, and that improving learning and measuring learning are compatible with each other. The weak positive correlation between factors 3 and 4 suggests that participants who believe assessment improves learning are not in strong agreement that that same assessment information is an indicator of school quality.

Table 4.10

Conceptions of assessment inter-correlations: economics and accounting teachers

Factor	1	2	3	4
1 Assessment describes learning		-.422	.533	.416
2 Assessment has weaknesses			-.320	-.039
3 Assessment improves learning				.180
4 Assessment for school accountability				

Extraction Method: Maximum Likelihood.
Rotation Method: Promax with Kaiser Normalization.

4.4 Conclusion

The factor analysis sought to investigate the conceptions of assessment of teachers of senior accounting and economics, and whether these conceptions are similar to those held by teachers reported by G.T.L. Brown (2006). While, overall, the fit statistics suggest the model should be rejected, CFA reproduced some aspects of G.T.L. Brown's model, in particular, the *improvement* and *school accountability* conceptions; however, the data showed a poor fit to G.T.L. Brown's *irrelevant* and *student accountability* factors. Four of the six factor inter-correlations were also similar. Agreements were inverse correlations between the improvement conception and the irrelevance conception, and between the irrelevance conception and the school accountability conception; and positive correlations between improvement and school accountability and school accountability and student accountability. The differences were an inverse correlation between irrelevance and student accountability and a correlation greater than one between improvement and student accountability.

Exploratory factor analysis revealed a four-factor model that reproduced similarities to G.T.L. Brown's (2006) model, but also had some significant differences. Similarities and differences are summarised in Table 4.11. The exploratory factor – *assessment describes learning* – contains all of the items that in G.T.L. Brown's model load to *assessment is valid* and *assessment describes learning*. Items relating to G.T.L. Brown's *improvement*, *student accountability* and *irrelevant* factors also load to Factor 1. These associations can be explained because the teachers in this study work almost exclusively at the senior level of secondary schools where they have a focus

on awarding internally-assessed achievement standards. If teachers award qualifications then the assessment needs to be considered valid and not unfair.

Factor 2 *assessment has weaknesses* has items that contributed to G.T.L. Brown's (2006) irrelevant factor. Notably, G.T.L. Brown's student accountability factor was not reproduced. Two items from this factor – *assessment places students into categories* and *assessment is assigning a grade or level to student work*, did not load to any factors in the EFA.

Factor 3 is essentially a combination of G.T.L. Brown's two first order factors of *assessment improves learning* and *assessment improves teaching*. Factor 4 has the same items as G.T.L. Brown's school accountability factor, but while these items loaded to the same factor, survey results did not show a very strong degree of agreement with this concept. For example, 40% of participants did not agree that assessment is an accurate indicator of school quality and 45% did not agree that assessment is a good way to evaluate a school.

Table 4.11

Comparison of items and factors of G.T.L Brown's (2006) model with factors and items for economics and accounting teachers.

<i>Factors and items for G.T.L Brown's (2006) TCoAIIIA</i>			<i>Factors and items for economics and accounting teachers</i>	
Second order factors	First order factors	Items loading to the factor	First order factors	Items loading to the factor
Assessment improves teaching and learning	Assessment is valid	Item 6 Assessment results are trustworthy Item 24 Assessment results can be depended on Item 15 Assessment results are consistent	Assessment describes learning	Item 20 Assessment determines if students meet qualifications standards Item 3 Assessment is a way to determine how much students have learned from teaching Item 24 Assessment results can be depended on Item 6 Assessment results are trustworthy Item 12 Assessment establishes what students have learned
	Assessment describes abilities	Item 12 Assessment establishes what students have learned Item 3 Assessment is a way to determine how much students have learned from teaching Item 21 Assessment measures students' higher order thinking skills		Item 22 Assessment helps students improve their learning Item 21 Assessment measures students' higher order thinking skills Item 16 Assessment is unfair to students Item 15 Assessment results are consistent
	Assessment improves teaching	Item 5 Assessment is integrated with teaching practice Item 14 Assessment information modifies ongoing teaching of students Item 23 Assessment allows different students to get different instruction		Item 17 Assessment results are filed & ignored
	Assessment improves learning	Item 22 Assessment helps students improve their learning Item 4 Assessment provides feedback to students about their performance Item 13 Assessment feeds back to students their learning needs	Assessment improves learning and teaching	Item 14 Assessment information modifies ongoing teaching of students Item 4 Assessment provides feedback to students about their performance Item 13 Assessment feeds back to students their learning needs Item 5 Assessment is integrated with teaching practice Item 23 Assessment allows different students to get different instruction

Factors and items for G.T.L Brown's (2006) TCoAIIIA

Factors and items for economics and accounting teachers

Second order factors	First order factors	Items loading to the factor	First order factors	Items loading to the factor
	Assessment makes students accountable	Item 2 Assessment places students in categories Item 11 Assessment is assigning a grade or level to students Item 20 Assessment determines if students meet qualifications standards		
	Assessment makes schools accountable	Item 1 Assessment provides information on how well schools are doing Item 10 Assessment is an accurate indicator of a school's quality Item 19 Assessment is a good way to evaluate a school	Assessment for school accountability	Item 1 Assessment provides information on how well schools are doing Item 10 Assessment is an accurate indicator of a school's quality Item 19 Assessment is a good way to evaluate a school

Factors and items for G.T.L Brown's (2006) TCoAIIIA

Factors and items for economics and accounting teachers

	Second order factors	First order factors	Items loading to the factor	First order factors	Items loading to the factor
Assessment is Irrelevant	Assessment is bad		Item 25 Assessment interferes with teaching Question 16 Assessment is unfair to students Item 7 Assessment forces teachers to teach in a way against their beliefs	Assessment has weaknesses	Item 7 Assessment forces teachers to teach in a way against their beliefs Item 27 Assessment is an imprecise process Item 25 Assessment interferes with teaching Item 9 Assessment results should be treated cautiously because of measurement error Item 8 Teachers conduct assessments but make little use of the results
	Assessment is ignored		Item 17 Assessment results are filed & ignored Item 8 Teachers conduct assessments but make little use of the results Item 26 Assessment has little impact on teaching		
	Assessment is inaccurate		Item 27 Assessment is an imprecise process Item 9 Assessment results should be treated cautiously because of measurement error Item 18 Teachers should take into account the error and imprecision in all assessment		

CHAPTER 5

Qualitative findings

This chapter discusses the findings from the analysis of the interviews. The first section discusses the findings that interviewees' beliefs and actions towards internal assessment for NCEA could be conceptualised along a continuum. The second section describes the five key themes that emerged from the analysis: assessment as pedagogy; assessment for educational outcomes; assessment as practice; assessment as motivator; and assessment is detrimental.

5.1 Gatekeepers and ushers

Analysis of the qualitative data suggests that teachers' beliefs about, and actions towards, internal assessment for NCEA may be conceptualised on a continuum from usher (door-opener) behaviour to gatekeeper behaviour. Usher behaviour was epitomised by maximising internal assessment for the benefit of students' achievement – that is using NCEA internal assessment to open doors to achievement. Gatekeeper behaviour, on the other hand, was treating internal assessment as an extension of external examinations. There seemed to be three dimensions to these actions and beliefs: relating to the conditions under which internal assessment was conducted, the types of assessment tasks that teachers asked students to complete, and teachers' general attitude towards internal assessment. These typologies are summarised in a matrix (Figure 5.1) but described as a continuum because not all individual teacher actions typified the extremes.

Table 5.1

Typology of teachers' beliefs and actions towards NCEA internal assessment

		Assessment conditions	Attitudes	Assessment tasks
Teachers' beliefs about conducting internal assessment for NCEA	Teacher (Usher)	<ul style="list-style-type: none"> • Open book • Formative feedback provided • Authenticity assured through ongoing monitoring • Resubmissions 	<ul style="list-style-type: none"> • Positive towards NCEA and internal assessment • Internal assessment should give students more opportunities to achieve 	<ul style="list-style-type: none"> • Original or large adaptations to exemplars • Adaptions to meet student needs • Portfolios • Contextualised
	Examiner (Gatekeeper)	<ul style="list-style-type: none"> • Exam-like conditions • No or minimal feedback • No resubmissions • Authenticity a driver and big concern 	<ul style="list-style-type: none"> • Negative towards NCEA and internal assessment • Internal assessment too easy 	<ul style="list-style-type: none"> • Minor adaptations to exemplars • Adaptions to ensure authenticity

The majority of participants ($n=15$) tended towards maximising the use of internal assessment for students, thus displaying usher behaviour. They used internal assessment to create opportunities for students to succeed. Their attitude towards internal assessment was positive, and they were positive about the opportunities that internal assessment affords to students to achieve. They created courses in which a majority of credits could be gained through internal assessment and favoured tasks and assessment conditions that maximised student success. They designed assessment tasks to appeal to students' interests, used a portfolio approach where students could receive ongoing feedback, and allowed reassessment opportunities. One participant described the process she used for an internal assessment:

I made a booklet that gives the data and the tasks. The booklet spells it out, a) do the calculations, b) do the graphs, c) identify the trends. Then I mark that, because there's no point moving on if that's not correct. So I mark the calculations, because if they're not correct there will be follow-through errors. But if they don't get it right the first time I use that as an opportunity for them to resubmit. (Tearlag)

Usher behaviour also included carrying out internal assessment in open-book conditions. While this could raise concerns about the authenticity of students' work, those participants who used open-book approached authenticity through ongoing monitoring of the students' work, by knowing students' general capabilities, and by having different topics for different students. As Sionn explained:

However if you break it up, have a checklist and ask them about it – where did you get that information from, can you explain why you have got this here? We let them choose (the topic). And then of course if they choose different things then that will lend itself to being their own work.

Ailsa believed students should be encouraged to collaborate and did not see this as a threat to authenticity:

If students go away at interval and find out from the others they have done something wrong – isn't that learning? I actively encourage the students to collaborate. I would rather teach the students about the ethics of plagiarising.

A minority ($n=7$) of interview participants' attitudes and actions towards internal assessment were more akin to the role of a gatekeeper. They tended to treat internal assessments as being similar to examinations rather than recognising different purposes, forms and pedagogical meanings associated with internal assessment. Authenticity of student work was a driver and major concern, and many gatekeeper actions around internal assessment were determined by authenticity concerns. For example, a rationale for making modifications to the TKI exemplars was not to accommodate students' needs but rather to prevent students accessing answers on the internet:

Our concern right from the beginning was that students could go onto the NZQA website and just get everything in advance, so it was really important we changed them (Isla)

So we modify all our internals – the answers are just there – on the web. (Ceanag)

Gatekeeper behaviour also included elaborate mechanisms for ensuring authenticity, such as students gathering data for research projects outside of class but bringing the data to class and writing up their reports in test-like conditions.

At Year 13 we're going to do the research-based internal, but we're making them bring the stuff in and they will do their report. So they can research it as much as they like, but they have to bring it in and then at a stage they can't bring any more in. (Ruairidh)

I try to avoid copying issues by doing it in test conditions. (Sorcha)

Furthermore, gatekeeper behaviour included giving very little feedback during the internal assessment process. Students were given practice internal assessments, which were imitations of the actual assessment. Feedback on the practice assessment was provided, but once students had completed the actual internal assessment, the grade was final:

We give them a practice internal, we give them exemplars, then we give them the real thing. (Ceanag)

Some would say to me – I've only made a small mistake can I fix it? And I've said no – you can't. To me – do you get a chance to fix it at the end of the year? (Ruairidh)

Participants were not necessarily dichotomous within the typology, but rather had a mix of beliefs and actions; in some cases their actions contradicted their beliefs. For example, Ceanag, who displayed mostly gatekeeper tendencies, expressed disappointment at not being able to conduct internal assessment with an open-book, portfolio approach:

So they do their investigation, they bring their data back and they do their write-up in a test situation. I don't like it because it really puts pressure on kids who don't do well in exams and I think one of the reasons for internal assessment was so that they could be done at home.

Others who described their internal assessment practice in terms of gatekeeping also described actions that supported a desire to meet students' interests through internal assessment. Sorcha described how she had introduced a Business Studies internally-assessed achievement standard because the students found it more interesting and that she used local businesses as case studies. Failbhe, who seemed quite negative towards

NCEA overall, described giving feedback and resubmission opportunities to ensure that students gained the internally-assessed standards.

Participants displaying mostly usher behaviour sometimes expressed more gatekeeper-like views. For example, Morag believed very little feedback on internal assessments was permitted:

Say a student got Not Achieved and they are working on a resubmit, we're not supposed to provide much in the way of feedback – if at all.

This included using a combination of approaches. Participants assessed some standards through portfolios and other standards through a combination of portfolios and formal in-class tests. Some standards were assessed solely through testing. As Mhairi explains:

There is a combination. Often they are done under some formal in-class time, and it depends on the assessment. The government policy one at Level 2 is a combination of in-class and out of class Market failure at Level 3 is in two different parts, one is a project they do on their own and the second part is a peer review where I have created a fake person and they have to analyse and write up what is right and if it was wrong what they would do to correct it. That's the bit they do in class One at Level 2 is going to be three tests – pretty much.

Eilidh, who generally described her internal assessment practice as usher behaviour, had actions seemingly contrary to her beliefs in that she used test-like conditions for some internal assessments. However, this was to ensure students completed the internal assessments and gained the credits, as she had found in previous years that students omitted to submit parts of a portfolio:

I can finish the test in maybe two or three classes, whereas research will take two to three weeks. And everyone will do the test because when they're sitting in the class, they can't say I am not going to do it. With research you are waiting for three weeks and then they haven't done it. So nothing is done and they don't get the credits.

All participants, whether they tended more towards gatekeeper or more towards usher behaviour appeared to be genuine and thorough in their administration of internal assessments; their differing views stemmed from their beliefs about being a professional assessor. Those with gatekeeper tendencies believed ensuring authenticity, giving little feedback, and allowing few resubmission opportunities

created validity and fulfilled their roles as guardians of the qualification. Participants who displayed more usher-like behaviours tended to believe that the ability to internally assess students was designed to enable greater student achievement in NCEA and actively used the system to achieve that goal. They expressed support for the NCEA and preference for it over the previous New Zealand qualification system and other international systems.

5.2 Thematic analysis of the interviews

Five overarching themes were developed from thematic analysis of the interviews.

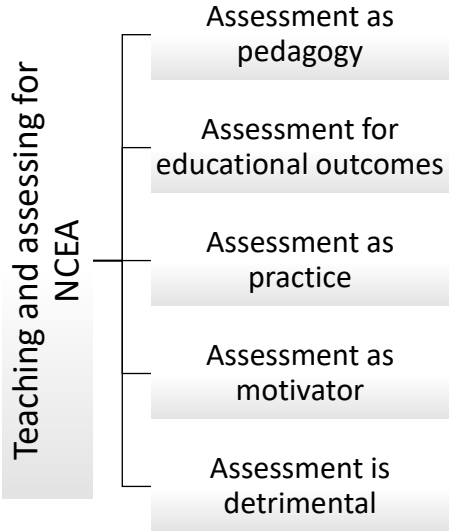


Figure 5.1. Overview of themes derived from the qualitative data

Table 5.2 gives a general description of each theme.

Table 5.2
Themes and their general description

Theme	Description
Assessment as pedagogy	Assessment is part of the teaching and learning process in the senior school.
Assessment for educational outcomes	Internal and external standards provide tools to enable students to achieve NCEA with the maximum possible grades, and mastery of subjects, and to prepare them for university, the workforce, and so on.
Assessment as practice	Assessment has an impact on teachers' practice. There are practical and professional issues with assessing for national qualifications.
Assessment as motivator	Assessment can motivate students and connect with their interests.
Assessment is detrimental	Assessment is detrimental to teaching and learning, and adds to teachers' workloads.

5.2.1 Assessment as pedagogy

All participants recognised that assessment, including NCEA internal assessment, is part of the learning and teaching process. The comments in this section show sound assessment literacy – that is, participants seemed aware of the various functions of assessment and the part they play in enacting these.

Assessment is for learning

All participants agreed in principle that a primary function of assessment is to ascertain students' learning. When asked what they thought the purpose of assessment is, most gave an answer reflecting a formative function. For example:

The purpose of assessment is to discover whether the students have learned anything and to judge what they've learned and to what extent they've learnt. (Scottiegh)

It's when students have the opportunity to show their learning – what they know and what they don't know. And they should be able to apply the concepts we've covered. (Catriona)

To provide feedback as to where they are at, where they are going, what they need to learn. (Ruairidh)

It has several purposes, it's a way of establishing if a student has reached a certain level of skill or knowledge, it's both formative and summative assessment. It's a signal to the teacher of how well the student is progressing and it's a signal to the student of how well they are progressing. (Malcolm)

Some added that assessment also served the function of evaluating their own teaching:

To inform your teaching and learning practice. To see whether the students are learning what you think they're learning, to see how effective your teaching programme is, to help identify areas they seem to be understanding well and areas you might need to cover again. It lets you know whether you're being effective or not. (Angus)

I think it is partly to see where the kids are at and the progress they are making and also you can reflect on how you're going as a teacher. (Maeve)

All gave examples of formative assessment in action that suggested this was very much a part of the teaching and learning process. They spoke of constantly checking students' learning, questioning them and providing feedback for the students' next learning steps. Some gave examples of how they track each student's learning and have mechanisms such as spreadsheets to document this:

My understanding of formative assessment is all that assessment that goes on right throughout a unit of work and it could be anything that is observed, general feedback from students, anything like that all contributes to formative assessment. (Maeve)

There's the informal assessment that you're constantly doing to see how the kids are going and whether you extend something or you move on because they've got the grasp of it. (Barra)

I ask pointed questions to various people and I know who is answering the questions. I keep track of who is answering questions and the sort of answers they give. (Seonaid)

Several gave examples of teaching the students to self-assess, whereby students ascertain their strengths and weaknesses by comparing their own work with criteria.

Mhairi specifically identified the increase in achievement this produced:

So they have to do a lot of evaluation of their performance. It has improved my results because I just used to do a bit of revision and some tests and didn't really get the time to evaluate from here ... They have to say these are the things they are good at and these are the things I was weak at. And it can't be more study, it needs to be specific. For example, label the graphs correctly or show the debits and credits etc., but it has to be really specific.

NCEA as part of teaching and learning

Many participants attempted to integrate NCEA internal assessment into the teaching and learning process, recognising the advantage to students of doing so. Examples were conducting assessment proximal to the teaching so the material is fresh in students' minds and having them complete internally-assessed standards as portfolios to enable ongoing feedback. Angus explained he uses several small assessment items or portfolios to alleviate anxiety and pressure to perform as a way to make assessment a more natural part of learning. He also commented that using this approach allows opportunities for feedback and for students to revisit learning. Sionn explained that by using an integrated approach he can provide feedback to students as they work and they can also reflect on their own progress:

I think it creates a lot of learning opportunities when you have that one-on-one conversation. It's not only about the assessment, it's about the relationship between myself and the student, and it's about getting the student to think or to think creatively or to sometimes question themselves. It allows them to reflect on what they've done or it allows them to reflect on their practices, the way they've approached things and maybe next time they should do it differently. So they (internal assessments) create a lot of valuable opportunities.

Those who used a portfolio approach also spoke of guiding the students through the task rather than expecting students to complete these independently. They emphasised students need scaffolding and without it often failed to submit anything. As Tearlag and Maeve explained:

If I left them to do that they would just fail. So I scaffold it. So one week we do xxx and I mark it over the weekend and anybody who needs to do it again can resubmit. Then we move on. (Tearlag)

I like to keep the timeframe tight. If you don't, too many will leave it until the night before it's due to do it. (Maeve)

Sorcha described how she breaks the teaching and assessment of internally-assessed standards into several assessable events, allowing resubmission if needed. Her description of the portfolio approach actually appeared to be teaching interspersed with a series of tests but did include a feedback loop with opportunities for students to learn through the assessment:

I'm doing a lot of portfolio things now, because you can do things separately. You can finish teaching something then do the assessment. You can give them a test, mark it and give the results then move on – rather than doing a big long assignment.

These actions show some participants were attempting to find synergies between the formative and summative functions of assessment and to use internal assessment as valuable learning opportunities.

5.2.2 Assessment for educational outcomes

Assessment for educational outcomes refers to teachers' rational strategies to enable students to gain qualifications with the highest possible grades as well as to prepare them for their future lives.

Assessment is for qualifications

All participants believed the ultimate aim of assessment in the senior secondary school is for students to achieve NCEA, with many stating that was the expectation of the students, their parents, and the school:

That's what they are here for! (Eilidh)

So they are prepared for the summative assessment. (Ruairidh)

But the overarching purpose, I guess, especially for Years 11-13 is working towards an NCEA qualification. (Morag)

The focus on NCEA in the senior secondary school became even more apparent when participants explained the differences between assessment in the senior secondary school and that in the junior school. Some described assessment in the junior school as formative and that feedback was to generate confidence and self-esteem. They emphasised the need to teach basic skills to the juniors and to continually revisit these until the students were competent:

At the junior level it's mainly formative and it's a signal to the students and the teacher of how well they are going. Maybe as an incentive, maybe to build their confidence. (Malcolm)

The purpose of assessment in the junior school is very much to build up their skills and in the senior school you tend to be concentrating on what's missing and the next steps but always with an eye on NCEA. (Tearlag)

Others described assessment in the junior school as preparation for the senior school and NCEA. Junior assessment mirrored NCEA with students being awarded the NCEA grades of Achieved, Merit, and Excellence for their work and students taught how to answer in the NCEA way of ‘explain’ and ‘fully explain’:

As a faculty we decide what we want our kids to be able to do in Year 11 and how can we make that happen? Things like ‘fully explain’ – are they learning how to ‘fully explain’ in Years 9 and 10? (Catriona)

We’re working it the same – we assess just like an achievement standard. We need to teach them the skills. That at Merit you can do this. They start economics at Year 9 so we may as well start assessing the way they are later. There’s lots of ways with 8 point scales we can mark the work. So we’ve written our mark schedules using those. (Doileag)

In addition to gaining qualifications, participants agreed they wanted to enable students to attain the best possible results including endorsements:

We have what’s called a scholar’s tie here, and so if you get over so many credits at excellence level you get a scholar’s tie, and they (the students) like that idea, so they strive to try – some of them do. (Barra)

Definitely, a huge push to get subject endorsements ... the motivation to get that external endorsement is huge. (Catriona)

At a decile 10 school our focus is on academic success so our main goal is to ensure that the students achieve ... and get the credits that they need for endorsement with excellence. (Doileag)

Strategies for educational outcomes

Participants identified a number of strategies they used to enhance students’ educational outcomes including managing the curriculum, increasing internal assessment, differentiating assessment, and students designing their own curriculum.

Managing the curriculum

A key strategy to achieve educational outcomes was to manage the curriculum by reducing the overall number of standards offered to students to enable them to achieve credits with endorsement rather than a high quantity of credits. Some reported that their schools had policies limiting the maximum number of credits permitted in each subject (often about 18 to 20), but participants themselves were the ultimate decision-makers over curriculum design. Schools with several classes of economics and/or accounting at the same level tended to have each class completing the same curriculum, so

participants spoke of negotiating with fellow teachers to determine which standards would be used. Mhairi summarises these thoughts and actions:

It is pointless to do quantity and not quality so I am all about quality. I want them to get the best grade they can, they don't need 24 credits, they need 14 to 18 credits and for Endorsement 14 credits. We don't need millions of credits, we want quality credits so give them enough time in the exam to write a quality answer and not be rushed.

In deciding which aspects of the curriculum to offer, participants tended to remove standards or topics that they thought students find challenging or uninteresting, or those that the teacher deems unnecessary. Nonetheless, their comments also reveal gender stereotyping and their own intuitive theories of motivation, rather than evidence-based decision-making. For example, many participants reported no longer teaching accounting concepts at Levels 2 (AS9117) and 3 (AS91404). As Kirsty explains:

Level 2 Concepts is boring and boys are not really concepts kind of people. I do things like inventory and MYOB as they are practical skills my students can use. Some don't go on to Level 3, they go and work in an office, so they need skills not concepts.

Barra agreed that boys, in general, did not like studying accounting concepts, but also thought this content was unnecessary until university level study:

I've dropped Concepts at Level 3, because the boys struggle with writing, most of the guys enjoy the numbers side, they find the concepts hard. I don't particularly like teaching it – I think that's more for university level accounting.

Angus agreed, although for slightly different reasons:

We've taken the Concepts standard out – the language is too difficult, especially for the international students. And it's too dry.

In economics, Ceanag and Scotaigh had removed AS91226 Analyse statistical data relating to two contemporary issues because it did not make sense to teach the statistics separately from the topic:

The students don't understand statistics, interms of the links between them and they get really confused about causes and effects, they just find it really hard. (Scotaigh)

Other participants explained that they still taught the whole curriculum but at some point they, or the students, decided whether to attempt the assessment. Participants either used their knowledge of students' achievement to counsel them about which external standards to attempt, or left students to make this decision themselves:

I'll teach the whole lot and be strategic for students. Individuals will look at where they're at and what they need and they might not attempt one in the final exam. Or I withdraw them from standards if they only need one or two standards. (Doileag)

I do the Trade external but they don't sit the exam. I withdraw them from it, but I want them to learn about the balance of payments, exchange rates and terms of trade so they've heard of them and got a grasp of them. (Failbhe)

Failbhe went on to explain that to gain Excellence in some standards students needed knowledge from other standards, so studying the whole curriculum was necessary, with which Kirsty agreed:

You need Trade to get Excellence because they can say how another aspect of the economy would change if the exchange rate was different. It means they can compare and contrast. (Failbhe)

And usually it's those extra layers (of knowledge from other topics) that help students get Excellence. (Kirsty)

Increasing internal assessment

A further strategy was to increase internal assessment and reduce the number of credits examined externally with a number claiming to offer up to 70% of available credits through internal assessment. All participants considered the internally-assessed standards more achievable and believed that students could gain higher grades because they are often given a reasonable timeframe in which to complete the internal assessment, they receive ongoing feedback, and teachers are able to offer resit opportunities:

So with internal assessment, even if they mess up, they still have a chance to resubmit in term two or term three. (Eilidh)

Internal assessment was also considered a more appropriate type of assessment for some students:

Internal assessment gives those students who don't memorise well the opportunity to pass something. (Siónn)

On a particular day how is that kid feeling? Maybe the exam is in the afternoon and in the morning they had another exam and they are tired. Maybe they're not feeling well, maybe something has happened in the family – all those factors affect the external. (Eilidh)

For some kids the external process just doesn't suit them – for whatever reason. Every year I'll have couple of kids who just don't do very well and you can't put your finger on it. (Ruairidh)

All participants agreed that students increased their likelihood of a higher overall course grade and more likelihood of success if they attempted fewer external standards. They claimed it was more advantageous to attempt only one or two standards in a three-hour examination rather than all three as it gave students more time to write in-depth answers. This strategy also increased students' ability to gain endorsements as endorsement favours fewer credits completed to a higher level. The three hour examination period was often perceived as not enough time to write three answers to Excellence level:

To get Excellence, which they really are striving for, they need more than an hour in the exam... we want quality credits so give them enough time in the exam so they can write a quality answer. (Mhairi)

I do two externals, they've only got three hours, and you can't compete if you do three with kids doing two. Ultimately doing one is better but then you don't cover enough of the course. (Faihbhe)

Some of my students last year in Level 3 said they spent two and a half hours doing just the company statements. NZQA have tried to make it so that each exam should take an hour but if a student wants to do it properly a standard takes more than an hour. (Kirsty)

Participants who taught economics had a further reason to increase internal assessment as some expressed dissatisfaction with the external assessment process:

There seems to be a real anomaly with the Level 3 externals, I've had students who are outstanding, but they only get one Excellence out of the three ... and last year the exam was awful ... you look at the papers and you can't figure out how they are marked. (Angus)

I think the external assessment in Economics is ruining the subject. Well you know everyone is saying how pedantic the markers are, and how it's turning students off the subject, it's really difficult to get Excellence, and now there is a huge amount of writing in the exams, so really high literacy requirements to pass. The graphing and maths side of it seem to have been diminished. But also the thought processes that have to go into writing three short essays in three hours is just too hard. (Aisla)

Differentiating assessment

Some participants described how they differentiated the curriculum and assessment to meet individual student needs and to ensure positive educational outcomes. Numerous techniques of differentiation were described such as multi-level study, extra tuition, blended learning techniques, school cluster groups (especially for Scholarship), teacher-created and proprietary resources:

I have made self-study units, both printed booklets and stuff online, for those more experienced kids. (Mhairi)

So I gave the students a taster (of the standard) and said if you want to sit this exam you need to come to tutorials ... about half chose to do that and half didn't. And that's fine. (Kirsty)

I have at least one (standard) up my sleeve just in case a Level 3 student needs it to go to uni but doesn't get the first one (internal standard). Maybe the resubmission didn't work out and so the student will ask to pick up something else. (Siann)

In addition, participants directed students to publicly available teaching resources such as YouTube, bought additional proprietary online and paper-based resources, and spoke of creating their own video teaching modules and posting these to the school intranet.

Student-designed curriculum

Participants reported that students, too, use NCEA standards to design courses with less assessment and decide whether to attempt internal and external achievement standards. Students too, are strategic in ensuring their own success and many students, once they have achieved the required number of credits, choose not to complete any further assessment:

We are noticing that students tailor their own courses and decide not to do 24 credits. They, too, want to do fewer credits and achieve better grades. Also, kids are very aware of where they are credit-wise and whether they need to sit any externals. (Morag)

Some kids think they've got enough credits and they don't even attempt the exams. There are always those that have no intention of getting endorsement. So usually they will pick one external and completely leave out the other one. (Sionn)

Two of my students just did the two micro papers and left the macro paper. Many of them go into externals having passed already, so they pick and choose which they'll do to get endorsement. (Seonaid)

As Barra explains, this is a rational decision on the students' part and Sionn believed students making their own choices is a skill to be encouraged:

It makes sense because say I'm offering 20 credits and the other 5 subjects offer 20 then that's 120 total. But how many do they have to get? We know it's 80, so there's 40 that they don't have to do. So is there any wonder that they're not going to do them all? (Barra)

To be honest I mean it's their learning and I think if they've got strategies around which to do and it works for them, they meet the requirements. I think it is a good skill isn't it? We may not like it as teachers but if it's an informed decision then I guess we should let them (Sionn)

Next steps in life

Participants were very aware of their responsibility to prepare students for their next steps in life – whatever they might be. While many focussed on preparing students for tertiary study, generally university, others realised this would not be the destination for all.

In terms of preparing students for university, some spoke of relationships they had with the local university and the expectations this imposed:

There is a strong relationship between us, Price Waterhouse and the uni – the girls have a direct career path in accounting. So, I talk to the uni before I change my accounting courses because the students have to meet what the uni wants. And Price Waterhouse gives scholarships to the students to study accounting at uni and they now have seven of my ex-students working for them (Ailsa)

When the kids leave here they go to university and I want them to be able to do the work. I want the uni to think – that student comes from my school and they know their stuff. (Barra)

My daughter has just finished four years at uni doing accounting and that's been really good feedback for me. Even in her stage three accounting she was still using some of the knowledge she learnt at school. So it was good to know that what we do in school is useful. (Maeve)

Some teachers also wanted students to understand the subject for the sake of it, not just to gain a qualification. Barra wanted his students to have a good understanding of economics, and Failbhe and Kirsty spoke of useful life skills:

I'm trying to prepare the kids for a life after school and their understanding in economics, so it's not just about getting over the goal line, but it's about making sure that they have a good understanding of what the subject is. (Barra)

Everybody needs to learn how to budget their groceries and if you understand the principles of accounting you understand so much more about every business, whether you're an employee or a boss. (Failbhe)

I've changed to teaching MYOB because I think it is a really practical skill that students can use, especially if they go and work in an office. (Kirsty)

This theme describes a summative view of assessment and the strategies teachers implement to ensure that students achieve NCEA with the highest possible grades. The strategies implemented and the reasons for these seem based on participants' intuitive theories about motivation and gender stereotypes, rather than an evidential base.

5.2.3 Assessment as practice

This theme reflects the major responsibility participants have in implementing and managing internal assessment for NCEA. Implementing NCEA internally-assessed standards has an impact on teachers' professional practice because teachers must choose assessment materials and the conditions under which the tasks are administered. The professional development offered to support NCEA also influences their practice as do the competing assessment regimes that have been established since the introduction of NCEA.

Implementing internally-assessed standards

Choosing tasks

Teachers have the choice to create their own NCEA internal assessment tasks, use the exemplars provided by the Ministry of Education (TKI exemplars), or buy commercial tasks. Most participants chose to use the TKI exemplars as a basis but to adapt them in some way; others said they created original tasks or used commercial materials with most using a combination of these approaches. They provided various reasons for adapting the TKI exemplars one of which seemed to be related to accountability in terms of meeting moderation requirements and also in terms of ensuring authenticity of student work. They also claimed they wanted to provide a relevant context for the assessment tasks, although their ideas about relevance were again linked to their own theories of motivation.

All participants agreed that they thought about what their students were interested in and tried to find suitable contexts for the internally-assessed tasks. As three said:

Well I definitely don't use the wedding planner exemplar at a boy's school!
(Doileag)

We're more of a farm school, so I do try and use agricultural examples.
(Kirsty)

Lots of the boys in this school have a background in the wine industry –
so I use that. (Barra)

Further examples of adapting for school context included simplifying the exemplars because, according to Scottaigh, “the exemplars on TKI are too verbose and too complicated”. Ailsa agreed stating “the exemplars on TKI are too complex, especially at Level 1.” Catriona wanted to “make sure they are kid friendly” and Tearlag did not think the exemplars suitable for students with English as a second language and “more than 50% of my classes are international students with poor to extremely poor English”. Wallace also stated that he was rarely happy with the exemplars and adapted them to appropriate reading levels and structure.

Three participants from lower decile schools said they made modifications to the exemplars because the exemplars make assumptions about students' access to resources such as the internet at home for research, access to bank accounts or the availability of business owners to interview. Their modifications included providing mock interviews, newspaper articles to analyse or bank accounts to examine rather than expecting students to provide these for themselves.

A further reason for modifying the TKI exemplars is the ease of access to the answers. Scotaigh sums up the problem saying, "If you use a TKI resource and you don't modify it the answers are on the internet and the kids can plagiarise". And for accounting, as Ailsa comments "you at least have to change the figures because the answers are on the web". Several participants reported that the National Moderators had recommended this approach at professional development days (Best Practice Workshops), saying original tasks were unlikely to meet moderation requirements.

Those who did create original tasks had the confidence and expertise to do so. Angus wrote most of his own as "the TKI ones have been riddled with problems from the start". He admitted to being one of the initial writers of exemplars saying "we didn't really know what we were doing back then", but the experience gave him the skills and confidence to write his own assessments. Other participants believed that internal assessment should be relevant to the students in the particular class and school, revealing their belief that if assessment is relevant, students are more likely to be engaged with it. Maeve thought students need to examine real, local businesses and therefore did not use other teachers' resources, Scotaigh believed economics should use contemporary events, and that "it was just as easy to write your own". He felt quite confident and spent time ensuring everything was right, but "if the moderator says I'm wrong, I'm wrong, but we were 100% on moderation last year". Seonaid thought that as long as the achievement standards were clearly written she can interpret them correctly and create her own assessment tasks:

I read the achievement standard, I read the schedule. It's all set out what you have to do for Achieved, Merit and Excellence, I don't need anyone else's schedule or task. I've never had to ask the moderator for advice and I get reasonably good moderator reports, so I feel confident I can do it myself.

Tearlag also explained the process for creating her own assessments which included reading the standard, the criteria, and the explanatory notes to make sure she understood what it all meant. She also referred to the Social Sciences curriculum, the Teaching and Learning guides, the conditions of assessment and the TKI exemplars. As she said, “I’m pulling stuff in from everywhere”.

Setting assessment conditions

The internally-assessed achievement standards do not necessarily specify the assessment conditions under which they must be completed, so teachers are able to decide these for themselves. Participants described a range of approaches such as closed-book, in-class tests, open-book but in-class tests, and open-book projects/portfolios. Individual participants did not subscribe to only one approach; they mixed these to suit the standards and to suit their own and the students’ workloads.

The rationale for conducting in-class tests varied, but was generally related to authenticity of student work, workload issues, and preparing students for end-of-year examinations.

We do the mall in class, so just all class tests, mostly for authenticity. (Isla)

It’s all done in the classroom without a computer. Each portfolio item takes an hour, or three hours over the week and the work doesn’t leave the classroom. That’s the only way I can control it. (Sorcha)

At the end of the year they go and sit exams so you have to prepare them properly. Why should we assess them differently? (Scotai gh)

The reality is, if you get the kids to do research it’s a lot of work as you have to do it yourself, but an assessment you can put together in a few hours and it’s much quicker to mark. (Ruairidh)

For some participants the in-class test was a one-off opportunity to achieve the standard, but others carried out several tests spread over a period of time:

If we do have tests, it’s a series so it’s more integrated with the teaching. It takes the anxiety away from them, this whole high stakes thing where you have to perform this particular day in this particular test. (Angus)

A further mode is for students to complete part of the assessment in their own time and the rest under test conditions. Students bring their data to class (e.g., interviews, statistics, newspaper articles) to write a report or answer questions:

So out of class they identify some government policies and the impacts on different areas, they come up with a sheet which they bring back to class and then they write it up formally in class. So, sort of a project leading to a closed-book submission. (Mhairi)

At Year 13 they do the research-based internal, but we're making them bring their stuff in and they will do a report in class. They can gather as much as they want and then for four lessons they will write about it and that will be sitting by themselves – not allowed to talk. (Ruairidh)

Teachers reported using open-book, ongoing assessments to provide formative feedback to students, with some managing authenticity by not allowing students to take their work out of class and others by using their knowledge of the students:

I do the Employment one over a number of periods in class, so the integrity of their work is not called into question, which it can be if they do it at home. And this way I see what they are working on and I give them feedback as they go. (Barra)

They do it at home and in class. We give them say, three or four classes to get started, not test conditions. I'm here to help them, they can discuss with each other, even share data collection. I break it up into bits, have a checklist and ask them about it. So, that works both ways. They get feedback and get on the right track and I know what it is actually their work. (Siónn)

The thing is with open book assessments it's a little bit worrying that students can cheat but it is up to the teacher to take in those drafts and know where the students are actually at, so when you are reading their final assessment you have an idea what level they should be. (Ailsa)

The majority of participants allowed students an opportunity to reattempt an internal standard if it was not achieved initially, but some seemed to offer minimal feedback in the resubmission process. They described allowing students one opportunity to correct minor errors within a limited timeframe:

You give the student the paper back, point them towards the problem they identify the problem and it should be fixed within a range of about 20 minutes. (Siónn)

I've got a feeling that if students are working on an internal there can't be much in the way of feedback, in particular if they are doing a resubmit. (Morag)

We don't actually mark the script so they can't see what is right or wrong, we say can you see an issue or a problem? (Doileag)

Others described how their practice towards resubmissions had changed:

I used to find their first effort wasn't up to much and they would just want resubmits, so what I do now is give them opportunities for feedback as they are working. I'm leading at the front, and there's going to be no reassessment, so you have to fix it now as we're doing it. (Barra)

I've tried to make it clear to students there will no resits, but they weren't happy so I've relented and we'll have a resit. But they will get three completely different questions. (Ruairidh)

Setting assessment conditions appear to be guided by a range of pragmatic decisions related to individual standards and workload issues, along with beliefs (that are possibly erroneous) about how much formative feedback they are permitted to give to students in relation to internal assessment.

Professional learning and support

The two main sources of professional learning and support for internal assessment were provision from NZQA and informal teacher networks.

NZQA provision

NZQA provides support to teachers through Best Practice Workshops and through moderators giving feedback on assessment tasks and marked work. NZQA also employs teachers as moderators and previous experience as moderators was cited as a key professional support. Angus explained that his previous experience as a moderator and exemplar writer provided him with skills to execute the internal assessment process effectively. Maeve and Isla also believed being a moderator enabled them to know what was required, and Malcolm thought previous experience as a moderator not only gave him confidence to produce his own materials but also to seek advice from the moderator if required. He considered that he had a good relationship with the moderator and would phone him for advice if necessary.

There were mixed responses regarding the usefulness of Best Practice Workshops and moderator feedback. Teachers appreciated the visibility of the National Moderator, as Sorcha says “I think she has done quite a good job because she is so visible. Previously, we didn’t always know who the moderator was”. Malcolm agreed that the workshops had been useful and had led to a change in the relationship between moderators and teachers. On the other hand, some questioned the ability of moderators to provide sound professional advice. As Ailsa said, “What they say goes, they run the best practice workshops so people have to listen to them. But I do sometimes wonder how they get that job”. Others agreed that the moderators did not necessarily have more expertise. Mhairi claimed, “Sometimes I think I have got more experience than the moderator in this area and I think I am right. Just because you have the name ‘moderator’ does that mean you know it all?”

Other reasons for not respecting professional advice from the moderation process included a perceived change in moderation expectations when the National Moderator changed:

Of course the moderator changes and so does the standard. (Barra)

You get changes in the moderator and they change what is seen as the national standard. (Isla)

Similarly, outcomes from attending Best Practice Workshops were not always positive:

I did go to one and I didn’t feel I got the answers. I like things to be black and white because I want to get things right. (Sorcha)

I wasn’t happy with them (the BPW). We had exemplars and none of us could agree on the grade – it just ended in a big argument. (Tearlag)

It did get frustrating because the moderator would just say go back to the standard, go back to the standard and not give definite answers. (Isla)

I go to all the Best Practice Workshops and listen, then take that away and do my own thing. I did ask a question once – can I do this? Is this acceptable? The moderator just said ‘look at the standard’. Well that’s not helpful, I’ve already done that. (Seonaid)

Teacher networks

All participants thought networking with other teachers was a valuable source of professional advice. They spoke of numerous ways they do this, such as formally through NZCETA meetings and informally through former colleagues and other economics and accounting teachers known to them or through online Google groups:

I belong to the local CETA group, so we just got together and wrote our own internal assessments. (Seonaid)

I collaborate with my ex-boss. I sent her my internals and she sends me hers. (Scottiegh)

We're lucky here as we have a network of really good teachers. There are a couple who are very good at writing assessments and so we all share. (Kirsty)

Even though I'm the only economics teacher here I don't feel like it because there is so much support out there. You just ask a question on the Google group and people start replying and sending you their stuff. (Eli dh)

A good thing about the Best Practice workshops was the collegiality and meeting other teachers. You get really good ideas, not necessarily from the moderator, but from the other teachers there. (Isla)

Competing assessment regimes

A further consideration for practice was the presence of several assessment regimes within their school revealing the complex nature of assessment in New Zealand secondary schools. These included NCEA, New Zealand Scholarship, International Baccalaureate and Cambridge International Exams:

This school has a dual pathway so the students get streamed at the end of Year 10 into either Cambridge or NCEA. Those doing Cambridge are the ones most likely to sit Scholarship, but the curriculum isn't the same, so we need to run separate scholarship classes. (Ceanag)

The school doesn't like to put the teachers into one or the other (NCEA or Cambridge) so we end up teaching both. (Scottiegh)

As IB is a two-year course I have to separate them for the first year and teach all the stuff that is unique to each and then in the second year (Year 13) they can combine as there is a lot of crossover. Although sometimes we take the mas two separate classes at Year 13 as well! (Barra)

New Zealand scholarship examinations cover a more extensive curriculum than Level 3 NCEA and teachers with students completing both have to make special provisions. These included extra tutorials and combining with nearby schools for scholarship classes.

5.2.4 Assessment as motivator

Participants shared many ways they believed they used NCEA assessment to motivate students and to connect with student interests. All participants wanted students to take and like their subject, but teaching an optional subject in the senior secondary school means first attracting students to those subjects. Participants thought that the flexibility of NCEA helps as it gives options to them and to the students. They also described various other ways in which they motivate and interest students, through field trips, enterprise studies programmes and connecting their courses to reality as much as possible.

Adjusting the curriculum to create interest

The flexibility of NCEA has allowed teachers to create courses that they think will appeal to their students. Participants did this through choosing what they consider the most interesting (and achievable) standards from within the same domain:

We used to do supply, demand and the market (standards) at Level 1, but the students were really bored, and I must admit these topics are not very engaging. So now we do demand, production and the circular flow and we've found the student engagement is way higher. (Kirsty)

We've changed from the statistics standard to employment, because the stats one just doesn't make sense and is boring. But everyone gets the idea that there are employed and unemployed people. (Ceanag)

We actually do a combination of achievement standards and unit standards for our Level 2 Business course because we want to appeal to a wide range of students, not just those who are academic. (Isla)

Participants also spoke of omitting standards that assess generic skills such as carrying out research projects or making decisions due to their repetition among subjects:

In accounting I've dropped the decision-making standards, because they do that in economics and they do that in a lot of the other subjects as well, and because they're going to get it here, here and here – they get bored with that. (Barra)

I don't think my kids are that turned on with big research projects – they do the min so many subjects and by the time they get to Year 13 they are over the m – so that's why I drop the m (Sorcha)

A further strategy used to motivate student interest is to mix domains. Most commonly, teachers of accounting and economics have introduced Business Studies standards into their courses. Accounting teachers saw the Business Studies Level 2 achievement standard 90848, Carry out, review and refine a business activity within a community context with guidance, as an authentic opportunity around which to base accounting:

We are looking at introducing the Business Standard (AS90848) because the students like it. They really like all that enterprise stuff – setting up a business, making things, having a market day and we can build the accounting around that. (Ailsa)

I used to have Young Enterprise as extra-curricula, but now we have the Business Studies standard, so I've put it in my course and we've based Inventory around it. (Maeve)

Other participants provided examples that were designed to interest and motivate students in their subjects:

I've put the human resources standard (from Business Studies) into Level 1 Accounting so the kids will see what happens about employing people and the other two Business internals for Level 1 are in our Economics course. It gives them a broader understanding. (Maeve)

We've developed a course called Commerce Studies and it has taken bits from Economics, Accounting, Legal Studies and Business Studies. It's totally internal at levels 1 and 2, so it appeals to a wide range of students. (Morag)

At level 1, the economics teacher and I teach a combo of economics and accounting – so the kids get an introduction to both. It's great because we can choose the best standards from each subject. And Social Studies do a Business standard at Level 2 for motivation. (Kirsty)

Achieving NCEA credits was also considered a motivating factor and two participants explained how they ensured students complete some internally-assessed standards early in Year 11 to provide them with the confidence that they can achieve. As one shared:

I do the internals first. Then the kids know they can do it, it increases their morale. They get an Achieved or Merit and can think "I'm not crap after all". (Eilidh)

Connecting with life

Participants also believed it was important to motivate interest in accounting and economics by connecting assessment tasks with students' interests and real life. The ability to write their own internally-assessed tasks and to adapt TKI exemplars allowed participants to contextualise assessment to local issues or students' interests. Catriona shared, "We definitely spend a lot of time thinking about our students and what they like to do and then rewrite the internals to suit". Barra also used examples that he thought appealed to his students, saying:

At Level one you've got to do a community organisation, so I choose a club they might know or have an affinity with, a surf-lifesaving club, a sports club, a football club or something they will know. A good thing is that you can tailor it to what the kids know.

All participants from Christchurch spoke of using the rebuild of the city after the devastating 2011 earthquake as a context for the employment topic in economics. As one said, "it would be silly to miss the opportunity the Christchurch rebuild is bringing for employment here". Two participants from Dunedin, had used the restructuring at the Invermay Agricultural Research Centre as a context for the employment topic and the Cadbury factory for studying producers. Scotaigh, who taught in a boy's school, had used the 2011 Rugby World Cup as a context to study employment, as he believed it interested the boys and was topical at the time.

Eilidh used a school community issue for Level 1 economics where students investigated the costs and benefits of making the school canteen food healthier. They then gave their research information to the school principal and "he actually changed the food, so the students did something real". Participants who teach accounting gave numerous examples of using local businesses as sites to study topics such as inventory, job costing or cash management. Using local supermarkets was popular because they closely monitor inventory; as Doileag said:

Their turnover is so high, and they know if there is an event at the local park in summer they will need to stock more cold drinks – the kids like that kind of stuff – it's so real. (Doileag)

I take my students to see a stock and feed business – this is real life Canterbury and I base the inventory internal around that. (Ailsa)

In Level 3 Accounting the students have to write a report for the Analysis standard – I let mine pick their own company, so it's a company that actually interests them. A couple of years ago a few did Pumpkin Patch and actually bought some of their shares. This year quite a few chose Xero and again want to buy some shares. So, I think that's pretty authentic learning. (Kirsty)

I'm hoping to use the local marae⁴ for the community organisation standard at level 1. Maybe they won't give me their exact financials but they could give some general figures that we can turn into a report. (Doileag)

Participants sometimes included field trips as part of the internal assessment process. Those who used supermarkets as a context to study inventory also included a visit to see this in action. Doileag, who worked in a boy's school, also took her students to Repco (a car parts and accessories retailer) as she believed this type of business appealed to the boys. Her plans to use the local marae as an example of accounting in a community organisation would include a visit to the marae.

Ailsa organises field trips to Wellington that include visiting the Treasury and the Reserve Bank for students at Level 3 to see economic policy in action, and at Level 1 she uses trips to local businesses for the production standard. Maeve tries to include visits to businesses, particularly those that belong to students' parents.

Creating businesses

Participants believed that having students create their own business was a further way to enhance students' motivation to study economics and accounting. This process can form part of an NCEA assessment programme, can be entered in various competitions offered by the Young Enterprise Scheme (YES), or can be informal. The Lion Foundation YES is a programme in which students set up a company, create their own product or service and sell it in a market. YES can be part of an NCEA qualification or offered as a co-curricular option. Participants described using Business Studies standards simultaneously with YES:

⁴ A marae is a communal and sacred meeting place which provides space for religious, educational, and community activities.

Young Enterprise works really well within accounting. Because whatever the kids come up with they have to do Inventory, so I do that with them for the internal. They've got to process financial information so that's where the MYOB processing internal can come in. (Maeve)

The Young Enterprise programme pretty much underpins all those standards in Year 13 accounting. (Mhairi)

Those who did not formally assess the students' businesses included them in their programmes or as co-curricular activities due to their perceived motivational impact. Sorcha was particularly supportive of YES but ran it outside the curriculum. However, she thought taking part could be life-changing for students and spoke of students who had changed career plans after successfully running businesses. Her students' successes in national competitions brought prestige to her subject and motivated them to study accounting and economics. She believed students were motivated by the competitive nature of the scheme as well as by the potential to make money, citing a student who had bought a car with her share of profits.

Angus said his students no longer entered YES competitions but they still created businesses and ran market days within their own school. He did not use NCEA standards to assess the businesses but gave students feedback on the process. He said he was always impressed by the students' confidence to order goods and outsource screen-printing and then sell their products either at school market days or on public internet markets such as Trademe. He also believed creating these opportunities for students increased their motivation for studying accounting and economics.

5.2.5 Assessment is detrimental

This theme encompasses views regarding possible negative consequences and undesirable effects of NCEA internal assessment, including concerns that NCEA internal assessment was not useful to the teaching and learning process and/or was damaging to teachers' working conditions and professionalism. Implementing internal assessment has had undesirable consequences for teachers.

Not useful to teaching and learning

Aspects of internal assessment that could be considered not helpful to teaching and learning included conducting internal assessment as testing which related closely to concerns about authenticity of student work. The reliability of internal assessment was questioned, and teacher participants expressed doubts about teaching students to be assessed and students being over-assessed.

Assessment as testing

Internal assessment for NCEA was seen by some participants as a form of testing and administered as an extension of the examinations. Many conducted some, if not all, internal assessment as tests, rather than allowing students to complete project-based tasks. Even formative processes in preparation for internal assessment included elements of testing:

At our school teachers are putting some of their internal assessment as part of the mid-year exams, so I'm certainly going to do that this year because it will reduce my marking load. (Scotai gh)

Even if it's open book we do it under test conditions – for authenticity. (Mhairi)

I've gone the way of an assessment – a one hour assessment, just like the end of the year. (Ruairi dh)

With the internals we try to do the formative assessment early, so if we need to, we can go back over things. (Isia)

For the internals what I usually do instead of formative assessment is give them quick quizzes throughout the teaching sequence and before an assessment I will always do a practice assessment. (Ceanag)

They evaluate their practice tests, what they have done well and what they need to work on before they sit the actual test. (Mhairi)

Those who taught in schools with more than one class at a year level spoke of all students completing the same or similar internal assessments, under the same conditions and on the same day, negating the ability for internal assessment to meet individual student needs:

It depends on the department but for our department we assess them the same because we think they are going to get assessed by the same assessment at the end of year so why should we assess them differently? (Scotai gh)

When it comes to internals they're written and moderated by one or two people and then they're sent out to all the teachers teaching that. There's no altering for specific types of learners and it's all done on the same day. You could be assessing in week 4 if your class finished early, but instead everyone assesses in week 6. I'd rather do things like presentations or debates or blogs, but instead it is paper-based tests. I think our kids could definitely be doing some varied kind of assessment. (Catriona)

Authenticity of student work

The authenticity of student work for internal assessment was a major concern for all participants, and they used varying approaches to ensure authenticity. For some, this concern led to implementing internal assessment as tests. Dealing with past incidences of plagiarism and cheating had made participants wary, but some also lamented that assessment had to be carried out this way. For others, authenticity was managed through closely monitoring students' work, rather than necessarily implementing test-like conditions:

We run them under test conditions – for authenticity purposes – we have had cheating trouble, so we've become mandatory on that, which is a bit sad. I'm sad we can't do it in a freer way. (Ceanag)

I try to avoid copying issues by doing it under test conditions. These girls could easily get a lot of help, it's a boarding house, so they could work together. So I don't put temptation in their way. And without computers there's no plagiarism either. So that's the way I deal with the whole issue. (Sorcha)

There's a real issue with kids sharing each other's work, they don't care. You repeatedly tell them not to do it, but they just ignore you. Last year – I had this kid who let another copy his work and I said "why did you let him do it?" and he said he felt sorry for him (Ruairidh)

Ceanag indicated that plagiarism had reached new levels due to students accessing the internet, saying:

Previously, when we didn't have the internet so freely available, you had to find something to copy, so you can't imagine how they could do that. Maybe you had mum and dad helping you, but good for them they were probably learning something too. But now, we have students paying for essays over the internet.

As a result, schools had created extensive mechanisms to prevent plagiarism or to catch culprits. Scottaigh described how his school had instituted Turnitin to detect plagiarism, and a number indicated they required students to sign authenticity statements and school management expected to know assessment conditions for all internally-assessed tasks. But, as Wallace describes, discovering non-authentic work is not without its own problems:

I realised one had copied from the other, had reordered paragraphs and changed sentences a bit. But finding that doesn't win me friends – the parents don't like it, it's a hassle for me, and it's time-consuming, so the easy thing would be for me to ignore it and I wonder if people do. I have to establish that they copied – which is not always easy. And then one had to own up, the whole thing took a while, and eventually one got Not Achieved.

Other participants approached things differently. Catriona describes how she teaches students to use the internet to find information and to paraphrase in their own words. She thought it important to teach them not to plagiarise at all:

I show them how to use Google, but also how to show they understand what they are talking about. I did an exemplar paragraph for one of my students. I googled the term and we found a nice meaning so I wrote that so they knew what to do and how to make it their own work ... It's important to teach the kids not to do that (plagiarise) because they can't get away with it at uni if that's where they're going. So why shouldn't we be carrying that down and saying we're giving you this opportunity to use the internet, to word process this report, do it in your own time and then hand it in? And with that comes responsibility and if you get caught this is what happens.

Barra and Sionn described how they closely monitored students as they completed internal assessment tasks, which served the purposes of providing formative feedback and ensuring authenticity of students' work:

For the Employment standard, I do that over a number of periods in class, the advantage of that is the integrity of their work is not brought into question. Whereas if they do it at home, that can be an issue. Although, even if they do bits at home I get them to show me what they are working on and I will give them feedback as they go. I'm not sure if that's the right thing to do, but I can see that they are working away. I've seen what they are doing so when they hand it in I know it's their work. (Barra)

In economics for example, we let them choose which industry they will investigate and then of course they all choose different things, so that lends itself to being their own work. But somebody could produce me some work and I could just ask them a couple of questions and I would know whether it's theirs or not. (Sionn)

Teaching them to be assessed

While participants did not say they actively taught towards the test, there were hints of NCEA assessment driving curriculum and pedagogy. Tearlag summed up a general feeling by saying “Even though they say you should teach as though NCEA wasn’t there, the reality is that you’ve always got your eye on the assessments and you’ve always got your eye on time”. Others agreed with comments such as:

We’re supposed to teach them to learn but sometimes I think we teach them to be assessed. (Scottiegh)

And probably I teach now more to assessment than I would have done when I first started teaching. I want them to get the results so stuff that is superfluous to the standard – a lot has gone which I feel sad about. (Doileag)

Eilidh did not believe she taught to the assessment, but the following comments suggests her feedback and instruction to the students is focused on achieving NCEA:

When I give them the assessment, the first thing I do is explain the questions, tell them to read the one by one and ask – what do you think this assessment is about? So they read the title and give some views. Then we read the first question and work out what this is about. And if they don’t get it – I can give them an example.

Other comments explained a focus on guiding students towards the higher grade levels of NCEA by explaining what they need to improve for Merit or Excellence. This included teaching the terminology of NCEA assessments:

We need to teach them the skills. That at Merit you can do this. (Sorcha)

I will give it (a formative assessment) back to the students with a model answer and I’ve taught them how to read the marking schedule. So we can really look at it and see where do I think they sit and what they need to do to get better. They’ll look at it and give themselves a grade, and then write three things down that they think they need to do next to bring this grade up. (Catriona)

When I do tests, formative tests in accounting I will break it up into ledgers, balance adjustments and statements. So I can say to the student you are really good at doing financial statements, but you need to work on getting ledgers to Excellence level. (Kirsty)

They need to understand what ‘in depth’ is and what’s going to give you the better grades. (Doileag)

I explain to them what ‘describe’ means, what ‘explain’ means and whenever they are doing exercises in class I just say that little thing, describe, explain. (Eilidh)

However, as Ruairidh and Catriona shared, it is an imperative of their job that students pass NCEA assessments:

I could sit in my classroom and talk about the economy all day and get the kids involved and have really interesting economic conversations, but that won’t help them pass at the end of the year. (Ruairidh)

What I find really hard is some of my students saying constantly from day one – is this what we need to do to get Excellence? I feel like saying can we just take a step back, let me teach you the basics, understand those then we’ll go deeper and then once you understand that you’ll be able to get Excellence. But they’re under the impression that here’s your model answer, memorise this and in the exam you’ll be fine. But then, if they understand it but don’t know how to write an NCEA answer they won’t pass.... So it’s a toss-up between teaching them to understand the content and then teaching them how to show that they understand the content. (Catriona)

Too much assessment

Some participants expressed concern about the amount of assessment required of students and students themselves made decisions about how much assessment they would complete. As some explained:

The students are in danger of being over-assessed because they’re continually doing internals. They’ve always got internals on the go and trying to balance the load between what they’re doing and what they’re not doing. (Scottiegh)

They can be under so much pressure. If you do six subjects and there’s two internals for each subject – that’s 12 internals. You have 40 weeks minus weeks for all the other things like exams, and term 4 is pretty much revision. So you have about 30 weeks to do 12 internals. So they’re having to deal with an internal every couple of weeks. Homework is just about a thing of the past as the students always have an internal they have to work on. (Wallace)

I think our kids are assessed so much that maybe we could just drop the exams and three internals or four internals would be enough. (Catriona)

As discussed in Section 5.2.2 on assessment for educational outcomes, a student may react to too much assessment by not attempting standards. While this practice, referred to as “opt-out”, was considered rational by some participants, others introduced mechanisms to limit it because students either risked not achieving enough credits or

not studying enough of the curriculum. Strategies to reduce opt-out included conducting internally-assessed achievement standards in class time to compel students to at least attempt these, reducing the number of credits overall, and completing internal assessments early in the school year:

The more credits you offer the more 'picky' the students can get – they don't do the externals because they have enough credits. (Kirsty)

When you're doing your internals towards the end of the year the students have already got the credits they need. Last year only about half handed the last one in. So this year I did the internals near the beginning. (Scottiegh)

Others also expressed concern about removing assessment standards from their programmes as this amounted to reducing the curriculum. Ceanag thought that this approach led to subjects being too compartmentalised, and students missed the big ideas and were not able to make connections between and among topics. She was also concerned that schools could make themselves look good in terms of NCEA results but thought “the reality is we're teaching a lot less,” and that while students could achieve more highly over fewer standards she wondered “whether that created a better educated student – I don't know”. Barra agreed a downside to not teaching the whole curriculum was teachers creating their own curriculum, stating “it becomes my decision as to what I think is important, which might be completely different to your decision”. Wallace was also concerned that students would miss key content, citing the example of excluding accounting concepts would deny students an understanding of the theoretical basis of accounting. Doileag agreed that not teaching accounting concepts was akin to “playing rugby without the rules”. And Tearlag thought student opt-out compounded a lack of curriculum coverage:

What is happening is that I will choose not to do all the standards, then the students will choose not to do some, so in the end they only have 14 credits – they haven't done much economics at all.

Reliability and validity

A number of comments were made that related to reliability and validity of internal assessment for NCEA. In terms of reliability, participants expressed concerns about how much help other teachers may give students to pass internal assessment, how much help they themselves were giving the students, and whether all students in New Zealand had equal opportunities to achieve:

There's no consistency, some people are making it easier for their kids, some people are making it harder for their kids. Everyone's marking is different. (Eilidh)

We have the situation where you have internals done at different schools under different conditions with different assessments. People are going to want their kids to do well. So, for example, I could offer the girls here resubmits until they give up. I've got 15 kids in that class and 15 in another class, so I can do that. But if I've got 37 kids, like when I was at my previous school and 37 kids in another class – can I really get them to resubmit, resubmit, resubmit? (Fai lbhe)

I give students one practice and an actual, whereas some schools give three practices then the internal. So – it's not fair on all the students. But there are some teachers who abuse the system to get their kids a better mark. (Kirsty)

At my previous school my perception was that the challenge is a lot harder than what is going on in other schools. We didn't give second chances – no resubmits. You sat the assessment and you got your mark. That was it. (Ruairidh)

Another problem with internals is that I don't know what other Economics teachers are doing, how much help are they giving their students? I don't know if other teachers give that specific feedback because there is pressure on teachers to get the student through. (Wallace)

Tearlag and Barra expressed concerns about the amount of feedback students should receive for an internally-assessed standard. Tearlag wondered if her scaffolding was in fact spoon-feeding, and Barra had concerns about at what point the work ceased to be that of the students and became his:

But when the kid comes in and wants a resubmission, and so I say "you've got this wrong can you explain it to me?" So, they sit there and can't, but if I ask them the right questions they probably could. But did the student know it or did they just give you answers to the questions that you asked? (Barra)

If they don't get it right the first time I use that as an opportunity for them to resubmit. So then the next point I have to scaffold them but sometimes it's a bit like spoon-feeding. (Tearlag)

The validity of teacher created assessments and exemplar assessments provided by the Ministry of Education was called into question by some:

To get Excellence a student must be able to describe the different weightings allocated to decisions. Then it comes back from the moderator – it would have been a stronger Merit if they had discussed that (the weightings). But it's not clear in the TKI task students have to do that, so I thought that's really dumb. (Mhairi)

I don't really understand what they wanted out of the TKI ones, or I don't think they do it well enough. (Sorcha)

Without exception, I've found the TKI ones need significant modification. They are poor examples of what students need to do to achieve the standard. (Ceanag)

Further threats to validity were the lack of exemplar tasks, exemplars of student answers for some standards, and unevenness in teachers' abilities to adapt exemplars appropriately:

Quite a few subjects didn't have exemplars of student work, but quite a few didn't have any tasks either which makes it so dependent on professional judgement. That can be difficult. (Isla)

Probably the one thing I do find frustrating is that we're all modifying them we're all doing the same. But I think that it's crazy that we're changing these tasks, and then NZQA expects that we will all get it right. I just can't see the logic in that. (Barra)

NCEA assessment is professionally damaging

All participants spoke of aspects of NCEA internal assessment that had negative consequences on their workload and their professionalism.

Assessment is a workload issue

All participants agreed that internal assessment required increased workload demands, including creating tasks, marking internally-assessed work, and monitoring compliance (e.g., sending samples for moderation):

I focus on creating one new assessment at a time, because I am the head of department and my department's really small. I am only getting one non-contact ... I teach all of the Economics so I can only work on one assessment for each level every year. (Eilidh)

I make up my own, but it takes up all of my holidays and then some ... it's so hard pulling all the information together, writing all the answers ... it takes so long to get it all done. (Tearlag)

One of the problems is time, teachers are often so damn busy that we just don't have the time to prepare exciting, new things. (Malcolm)

Marking and providing feedback is also time consuming:

The internals have added a lot of work for teachers in terms of the marking load. I'm reserving judgement on the portfolio approach, I think it's better for the students but worse for the teachers. (Scotaigh)

Scotaigh's reasons were keeping track of resubmissions, parts of portfolios being handed in at different times, keeping consistency in marking when it was spread over a period of time and dealing with absentee students. Tearlag and Doileag also explained the time-consuming nature of marking:

I'll mark over the weekend and anybody who needs to can resubmit. But this approach takes up all your time – it means you have assessments running through lunch time and after school. Sometimes I have study groups on a Saturday – anything like that. Private tutorial sessions – and that's what it's become. (Tearlag)

That's the trouble. When people think this is what I do, they're thinking five to ten kids and I've got 46 – that's 46 internals to mark each time. I mean it's fantastic because you get a class that has diverse personalities and all those things. But the workload is quite big. (Doileag)

Both Scotaigh and Isla expressed a preference for offering more external assessment as it entails less work for teachers, but their schools had policies to which they had to adhere and, as Isla said, the internal assessment topics tend to be more authentic learning:

It's one of those difficult things because I have to admit, that I would much rather drop an internal – the interpretation, company analysis one, the marking of it takes almost a week of the holidays. So if I was to drop one I'd rather drop that and do concepts. But in fact that is probably the most authentic real life thing that we cover, looking at a real company. (Isla)

And as Ruairidh explains, workload considerations are partly why he conducts internal assessment as a series of tests:

From a purely selfish point of view what I do is easier on my time. At my last school I used to work really long hours ... But my home life was suffering ... I was new and I had to create a lot. It's easier now, but I'm here at 7a.m every day and I'm still here at 4p.m you do your co-curricula, and still working on Sundays. So the reality is if the kids do research it's a lot of work as you have to do it yourself, but an assessment (test) you can put together in a few hours and it's much quicker to mark. (Ruairidh)

At the time of the interviews, a major realignment of the NCEA standards with the NZC had occurred and had exacerbated participants' workload. Several participants suggested that the increased internal assessment may have been motivated at least in part to reduce costs for NZQA at the expense of teachers' workloads:

The realignment has brought new standards and it's given us the opportunity to change some of the standards we were doing. But it is a workload issue to change topics too much and develop new assessments. (Angus)

There has been a big change with the realignment and little recognition in terms of the impact that has had. To all of a sudden go from 25 – 30% internally assessed to 50% internally assessed, the workload has gone up and there has been little recognition from the union and from NZQA. I'm probably being cynical but I think one of the biggest reasons (for more internal assessment) is that exams are very expensive for NZQA and if they don't have to employ as many markers and examiners then they save money. (Mhairi)

But us doing all that extra internal assessment on the cheap is pretty annoying, all the extra moderation – that is quite a lot of extra worry and work for no extra pay. (Sorcha)

Mhairi added that adding one more internal per subject may not seem like much, but as she explains:

What has been the impact on courses and workload with at least one more internal? It's pretty much the same content, but the process of assessing an internal is much longer than assessing an external. You can teach the stuff for an external in about four to five weeks and you will have a couple of tests and that's it. But for an internal you might teach the stuff for four to five weeks and it might take you another two or three weeks to get through that whole process of assessment and re-submission, and then the workload to mark it all. (Mhairi)

Moderation damages professionalism

Internal assessment allows teachers to award credits towards national qualifications so NZQA monitor the process and teachers are held accountable for their decisions. For a number of participants this system has been professionally damaging with many using emotive language such as “hammered” or “slammed” by the moderator, or being “hailed over the coals” and “given a big stick” by school management.

A key issue is the reaction of school management to moderation reports. There seems to be an expectation from school management that teachers will receive one hundred percent agreement on their awarding of grades and require few, if any, modifications to the tasks they have designed:

You want your 8/8, the senior management and schools want 8/8. But I guess my philosophy is that there is always going to be a difference – so as long as you are close to 8/8 that would be fine. (Ruairi dh)

Scotaigh indicated that he was confident but he spent time “making sure the marking is right”, because “if you’re wrong you get hauled over the coals ... the school wants to know what is going on”. He added “I suppose management could argue we’re professional teachers and should be getting that right but I thought we were allowed to make mistakes as well”.

This sentiment was echoed by a number of others where Tearlag was “petrified it’s (moderation) going to fail miserably” and Sorcha shared that “you can get into a bit of trouble if you don’t get it right”. Catriona also explained that her department had “one terrible moderation. I think it was 0 out 8 ... we definitely had to explain that to our head of faculty and show him what we had done to fix it”. And Kirsty felt “you are judged by your moderation”.

Some considered themselves to be equally, if not more, qualified than the current moderator and disagreed with moderator decisions:

Sometimes I think I have more experience than the moderator in this area and I think I am right. But I still have to justify it (to school management). (Mhairi)

I had one last year where I only got 2/8, so I was a bit worried about that this year. Because usually I have 8/8. I did some facilitating when NCEA came in – so I was pretty onto it ... I don't necessarily think she was right. I think I did what she said I was supposed to do. I'd sent it to a teacher in another school and he didn't fault my judgements – so he must be wrong as well! We can't have it going back to our schools looking like there is a fault with us. It's a professional thing. (Sorcha).

Isla and Failbhe explained the issue of changing moderators and standards:

When there was a change in moderator there was a major change in expectations and suddenly the word 'describe', actually means 'explain' and that was huge, and a lot of people got slammed on that one. The problem with it is the reaction of senior management, but when you explain the new moderator decided that they want this instead of that, you're okay but it is frustrating. (Isla)

The year before, the year before, the year before 100% agreement, 100% agreement, 100% agreement. But last year 0 out of 8, all of them were Excellence under the old system but none of them had introduced an additional piece of information. They were all Merit, but the school goes that's not great. (Failbhe)

Eilidh described a similar situation where she had used an assessment activity designed by another teacher which had been approved by the moderator, but her moderation report recommended changes:

The assessment was moderated already, someone else was teaching it, he had it moderated three times, nothing, no complaints, nothing. So that's why I adopted it. And I got that I had to change a question. Maybe the moderator was different? But I was confused.

Ensuring professional safety was a key mechanism in determining participants' decisions to use TKI exemplar materials rather than creating original assessments:

It is risky for moderation if we make too many changes or write our own. (Ailsa)

I tend to use TKI, modify it a small bit – so change the context. The reason I don't go for developing my own is that if you stray too far then you get hammered from the moderator, and the school management get on your back. (Barra)

According to Mhairi the national moderator encouraged such practice and had informed teachers to expect alterations to their original assessments and that it could take up two or three iterations before they would meet the standard. However, designing original materials that are then critiqued by the moderators leaves teachers vulnerable to school managers' adverse reactions:

He was saying you have to expect that there is going to be revisions required but you have to explain to your school why you failed moderation. He's saying it shouldn't be about failing or passing, it's about making it better for next time. But you have this difference of opinion or ideas of what moderation is about.

Past experience had left Tearlag fearful:

It was terrible, it was terrible until more recently. But even then I don't know if I've had enough rounds of moderation to get my confidence back. Because they were pinging us on the actual assessment and we'd just downloaded those from TKI – so we were being criticised for something we didn't write but we were told to use.

Pressure to perform

In addition to expectations with regards moderation agreement rates school managements also had expectations with regard to grades and pass rates that teachers and students would achieve. Participants voiced negative consequences that occurred if these expectations were not met:

Our faculty goals are at level one to have 45% of students achieving Merit or Excellence. Then level two I think is to have 80% Achieved. (Catriona)

They expect everyone to pass and at level one – 20-25% Excellence, at level two and three – 10% minimum Excellence. There is a report written for the Board of Trustees and when the results come out they are compared against the rest of New Zealand. And because we are decile 10 we're expected to be better than the average. (Ruairidh)

You get knifed in the back by your own school for having Merits, not Excellence ... you're being judged on your results. (Faihbhe)

Here, at level one, the results were very, very good, at level two they were not too bad, and at level three – they were poor. That's why I'm here. The other teacher is gone ... Same at my other school – you didn't actually lose your job, but there are huge expectations that you will perform (Ruairidh)

Participants also expressed an accountability to their students and internalised the students' pressure to perform.

I've taught them what to do and then they don't pass. They could say it's my teaching, I didn't teach them properly. So I find that reasonably difficult. (Barra)

But you do feel the pressure if they haven't done as well as they should have and another teacher's class has done better. (Catriona)

You've got to look at the subject next door. And if Music is generating 10 excellences out of 15 kids and you are generating one then what would you do as a student, which way would you go? (Fai lbhe)

We spend more time on stuff so they have a better chance of getting good marks. It's very competitive here, all the schools are good but there's not enough kids to go round. So then you get the issue of which school has the highest marks. (Tearlag)

5.3 Conclusion

This chapter has described the five key themes that were derived from the qualitative interviews. The themes indicate conceptions or beliefs participants have about NCEA internal assessment. These are: assessment as pedagogy; assessment for educational outcomes; assessment as practice; assessment as motivator; and assessment is detrimental. Participants shared many positive outcomes of assessing for NCEA. They discussed how internal assessment is an integral aspect of their pedagogy and that they considered it part of the teaching and learning process. They managed internal assessment to enhance educational outcomes for their students because success in NCEA is an expectation from students, their parents, school management and society at large. Participants recognised the potential motivational impact of internal assessment and strove to connect assessment with student interests and designed programmes and tasks to engage students in learning. However, some participants also believed that implementing internal assessment could be damaging to their professional reputations and decisions with regards the types of tasks used were made in light of this potential damage. Being an assessor of national qualifications requires accountability and responsibility and these were key drivers in determining participants' approach to internal assessment.

CHAPTER 6

Discussion of the strands

The purpose of this mixed methods explanatory sequential design was to investigate the conceptions of assessment of teachers of economics and accounting, how they assess the internally-assessed achievement standards for NCEA and the reasons for the decisions they make. Specific questions were:

1. What are the conceptions of assessment for teachers of accounting and economics?
2. What are the reasons for teachers of accounting and economics making the decisions they do about the internal assessment tasks they offer students?
3. What is the impact of their beliefs about assessment on their approaches to internal assessment design?

A goal of mixed methods research is to integrate effectively the data from both strands of the research to provide greater understanding of the phenomenon under investigation (Teddlie & Tashakkori, 2009). Fetters, Curry, and Creswell (2013) describe three approaches to integrating the data: integrating through narrative; integrating through data transformation; and integrating through joint displays. This thesis will primarily integrate the data through narrative with some use of joint displays. Within a narrative approach, the two data strands can be reported contiguously and/or through a weaving approach. Firstly, this chapter will report the two strands contiguously in order to answer the first two research questions. Chapter 7 will consider how the data from both methods connect and converge (Creswell, 2012).

6.1 Discussion of the quantitative findings

6.1.1 Confirmatory factor analysis

The quantitative aspect of this study explored economics and accounting teachers' conceptions of assessment using G.T.L. Brown's (2006) TCoAIIIA Inventory: CFA was used to evaluate the fit of the data to G.T.L. Brown's model. The analysis reproduced some aspects of the model, in particular, the *improvement* and *school accountability* conceptions of assessment. However, the data showed a poor fit to the *irrelevant* and *student accountability* conceptions and, overall, the fit statistics suggest that G.T.L. Brown model does not adequately fit the present data. The reproduction of the *improvement* conception shows that teachers in this study agree that assessment information is used in a formative way to improve student learning and to modify teaching. This finding supports the view that despite having a strong focus on external qualifications, participants still reflect the emphasis on *assessment for learning* in New Zealand's educational culture (G.T.L. Brown, 2011a; Ministry of Education, 2007, 2011).

It is not surprising that the *school accountability* conception of assessment was also reproduced, given the competition among schools that was established from 1990 onwards. Secondary schools compete for students and qualification pass rates are used as a marketing tool to attract them (Wylie, 2012), with those gaining higher pass rates considered better schools, which has a positive impact on roll numbers (Wylie & Bonne, 2016). NCEA pass rates are publicised in national newspapers (e.g., Dougan, 2016), making secondary schools publicly accountable for their assessment results. The Better Public Service Target (New Zealand Government, 2012), which introduced a target that 85% of all young people attain NCEA level 2 by 2017, has reinforced the concept of using NCEA results as an accountability tool for the education system.

It could be predicted that New Zealand secondary teachers would have expectations of *student accountability* and thus reproduce this factor. However, two of the three items in this factor (Item 2 – *Assessment places students into categories* and Item 11 – *Assessment is assigning students grades*) loaded weakly (.11 and .14 respectively), whereas G.T.L. Brown's (2011a) research on a more general sample of primary and secondary teachers reported these items loading strongly (.67 and .64 respectively).

The weak loadings reflect similar international research into secondary teachers' conceptions of assessment; for example, Gebril and Brown (2013) reported .32 and .15 loadings respectively for Egyptian teachers and Brown et al. (2009) reported .24 and .23 loadings for teachers in Hong Kong. It is likely that secondary teachers do hold a *student accountability* conception of assessment but that the items used by G.T.L. Brown (2006) to measure this conception are not appropriate in this context. Participants were in general agreement with all three items, that is, they agreed that *assessment places students into categories*, *assessment is assigning a grade to students* and *assessment determines if students meet qualifications standards*, but these beliefs do not cohere into a construct. It is unsurprising that participants agreed with these statements because, for teachers assessing senior students for NCEA qualifications, these statements are essentially facts. They are functions of assessment that participants in this research routinely carry out.

The *assessment is irrelevant* factor in G.T.L. Brown's (2006) model consists of three first-order factors: *assessment is bad*; *assessment is ignored*; *assessment is inaccurate*. While the CFA showed strong loadings of the items to the *assessment is bad* and *assessment is inaccurate* first order factors, the items contributing to the *assessment is ignored* factor (*assessment is filed and ignored*, *of little use* and *has little impact*) all loaded weakly. This finding has similarities to research carried out with secondary teachers in high-stakes assessment environments, such as Hong Kong, Queensland and Egypt where some items did not load well to the *assessment is irrelevant* first order factor (Brown et al., 2009; Brown & Lake, 2006; Gebril & Brown, 2013). Responses to individual items contributing to this factor also show general disagreement with several aspects of the *assessment is irrelevant* factor. For example, there was general disagreement that *assessment interferes with teaching*, that *assessment is unfair to students*, that *assessment results are filed and ignored*, that *teachers conduct assessments but make little use of the results* and that *assessment has little impact on teaching*. These disagreements, and weak loadings, are possibly to be expected for teachers in this study, as well as those in Hong Kong and Queensland, because for all three groups of teachers, assessment results (internal assessment in particular) have substantial utility because they are required to award qualifications. Another explanation is that this finding may also reflect the *assessment for learning* culture in New Zealand schools (G.T.L. Brown, 2011a; Ministry of Education, 2007, 2011); not

only does internal assessment for NCEA have utility for qualifications, it can also contribute to student learning.

The inter-correlations between the conceptions specified in the CFA also differ from G.T.L. Brown (2006) and align more with findings from international examination jurisdictions. The CFA revealed a highly positive correlation between the *student accountability* and *improvement* conceptions, a finding similar to the results of research carried out in Hong Kong ($r=.91$; Brown et al., 2009), China ($r=.80$; Brown et al., 2011) and Egypt ($r=.95$; Gebril & Brown, 2013). In contrast, G.T.L. Brown's (2011a) research with New Zealand primary and secondary teachers found a relatively weak correlation between these conceptions ($r=.23$). G.T.L. Brown (2011a) attributed the weak correlation to an *assessment for learning* culture in New Zealand schools, whereas the high correlation shown for other jurisdictions could be attributed to teachers expecting students to be more accountable for their own learning. Brown et al. (2009) concluded that the Hong Kong examination system has a large impact in that Hong Kong teachers consider making students accountable as a means of improving their learning, notions with which New Zealand accounting and economics teachers apparently agreed. A further explanation could be that teachers consider improved assessment results to be an *indication* of improved learning. These explanations would be consistent with teachers expecting senior secondary students to be more responsible for their own performance than younger students, and with secondary teachers more typically using grades to recognise learning. The strong correlation between improvement and student accountability is also consistent with Dann's (2014) explanation of assessment *as* learning, in which Dann promotes the prospect of assessment enabling students to understand and gain agency in their own learning. The structure of NCEA certainly encourages student accountability by allowing students to influence their own learning and assessment pathways (Absolum, Flockton, Hattie, Hipkins, & Reid, 2009; Nusche et al., 2012).

The CFA also established an inverse correlation between the *irrelevant* and *student accountability* conceptions of assessment, which suggests that utilising assessment results for student accountability is considered a relevant use of assessment by the teachers in this study. This aligns with findings from Hong Kong (Brown et al., 2009), which similarly found an inverse correlation and concluded that the Hong Kong

examination system has a large impact on beliefs about the relationship between student accountability and the relevance of assessment. The findings contrast with positive correlations between these conceptions for teachers not necessarily involved in preparing students for qualifications (G.T.L. Brown 2006, 2011a). Taken together, these results suggest that teachers tend to believe that use of assessment for student accountability is more relevant when those results are used for qualifications.

The participants in this study are distinctive in that their main, sometimes sole, focus is teaching in the senior secondary school and awarding NCEA qualifications which could explain why their conceptions of assessment differ from those of participants in previous New Zealand studies. G.T.L. Brown (2011a) compared New Zealand secondary and primary teachers and concluded that “the only statistically significant difference was the mean score for the student accountability conception which was more strongly endorsed by secondary teachers” (p. 45) which he attributed to their role in awarding qualifications. However, G.T.L. Brown’s report does not indicate whether or not secondary teachers in his study were solely focused on NCEA. Secondary teachers, when thinking about Year 9 and 10 students, may have conceptions of assessment more similar to those of primary teachers, because these students are generally not assessed for NCEA.

6.1.2 Exploratory factor analysis

A subsequent EFA resulted in a four-factor model which, in part, agrees with G.T.L. Brown’s (2006) previous work, but includes differences that appear to be linked to the role of teachers in assessing for qualifications, and differing from those identified by G.T.L. Brown. Factor1 *assessment describes learning* relates to a summative conception of assessment whereby assessment determines what students have learned and whether they have met required standards, but also includes items that relate more to the improvement of learning as well as the validity and usefulness of assessment. Similarities with G.T.L. Brown are that all the items his model associated with *assessment describes abilities* and *assessment is valid* are present in this factor. The association of validity concepts with summative concepts of assessment can be explained: assessment must be valid in order to perform a summative function for the award of qualifications.

In G.T.L. Brown's (2006) model the item *assessment determines if students meet qualifications standards* is associated with a student accountability conception. However, for the teachers in this study, meeting qualifications standards is part of the summative process. The item *assessment helps students improve their learning* is associated with an improvement conception in G.T.L. Brown, and its association with more summative functions of assessment for participants in this study may relate to the idea that teachers can make formative use of summative tests (Black et al., 2003). Black et al. describe situations where tests are opportunities to improve students' learning, because students will learn material to succeed in the tests. In addition, once tests are completed, feedback can be provided on the tests, and so formative use is made of summative assessment.

Two items (*assessment results are filed and ignored* and *assessment is unfair to students*) are associated with G.T.L. Brown's (2006) *assessment is irrelevant* factor. Participants in this study were in general disagreement with both statements, signalling that they regarded assessment results as neither ignored nor unfair. Their association with a summative conception of assessment is logical for teachers in this study because if assessment has a valuable summative function then the results are not filed and ignored and neither is assessment unfair to students. This indicates concurrence with Sadler's (1989) view that summative assessment is often geared to certification and this function of assessment has value for students and teachers (Madaus & Russell, 2011).

The association of these ten items into one factor could relate to the role participants have in assessing internally-assessed standards for NCEA because assessment for internally-assessed achievement standards can involve providing feedback as part of the summative process. A hallmark of the summative use of assessment is *determining if students meet qualifications standards*, and the co-dimensionality of this item with *assessment helps students improve their learning* could describe actions that relate to assessing NCEA internally-assessed achievement standards. Teachers awarding internally-assessed achievement standards do determine whether students meet qualifications standards, and may consider that this process helps students' learning. The co-dimensionality of these items could indicate a belief that internal assessment for NCEA can be used for formative purposes, despite its role in the awarding of

qualifications, which again is consistent with the assessment for learning culture prevalent in New Zealand schools (G.T.L. Brown, 2011a; Ministry of Education, 2007, 2011). It could also be consistent with teachers believing that formative and summative functions of assessment can be compatible and not necessarily “two isolated and completely different functions” (Black & Wiliam, 1998, p. 35). Teachers in this study may be developing synergies between the two functions of assessment, as Maxwell (2004) and Harlen (2006) claimed is possible. On the other hand, this co-dimensionality may signal that teachers are providing feedback to students to enable them to pass assessments rather than to improve their learning. This would support Torrance’s (2007) view that compliance with assessment criteria can be a proxy for actual improvement of learning. Because the teachers in this study are concerned with assessment contributing to the award of qualifications, they arguably have a strong motive to take this approach.

Although Factor 2 aligns with G.T.L. Brown’s (2006) *assessment is irrelevant* conception, the responses to the items contributing to this factor showed general disagreement or mixed responses to these negative aspects of assessment. The strong disagreement with the item *teachers conduct assessments but make little use of the results* is unsurprising because the accounting and economics teachers who participated in this study teach almost exclusively at the senior level of secondary school where there is a strong focus on assessing for qualifications, so they would obviously consider that the results are used. Because assessment results are required to award students’ qualifications, they are viewed as highly relevant and not ignored. This disagreement could also indicate consistency with the *assessment for learning* culture prevalent in New Zealand schools (G.T.L. Brown, 2011a; Ministry of Education, 2007, 2011); assessment results are needed to improve student learning, thus the results are useful. The mixed responses to *assessment forces teachers to teach in a way against their beliefs*, *assessment interferes with teaching*, and *assessment results should be treated cautiously because of measurement error* indicates varying views of the relationship between assessment and teaching. They could also indicate varying views with regard to the impact of assessing for internally-assessed standards on their teaching practice and the different practices teachers adopt towards internal assessment. This proposition is certainly supported by the limited literature on teacher practices for NCEA internal assessment whereby teachers have been reported as

conducting internal assessments as separate tests (East, 2016; Hume, 2006; Moeed, 2010). The practice of separating assessment into individual events could be seen to interfere with teaching and possibly forcing teachers to teach against their beliefs. On the other hand, East reported on teachers who provided multiple and ongoing opportunities over the year to attain the requirements of internally-assessed standards. Teachers in this study who adopted the latter stance could consider that conducting NCEA internal assessment in a progressive and more integrated way does not interfere with teaching.

Factor 3, *assessment for improvement*, is similar to two of G.T.L. Brown's (2006) first order factors *assessment improves teaching* and *assessment improves learning*, with one exception: the item *assessment helps students improve their learning* associated with G.T.L. Brown's conception did not load to this factor. Participants in this study did not seem to separate the *assessment improves learning* conception from the *assessment improves teaching* conception, rather, seeing them as integral to each other. This would align with Black et al.'s (2002) view that for formative assessment to occur the evidence gathered during an assessment process must be used to adapt teaching to meet learning needs. The agreement from participants to the items *assessment modifies the ongoing teaching of students*, *assessment is integrated with teaching practice*, *assessment provides feedback to students about their performance* and *assessment feeds back to students their learning needs* supports the notion that participants are adapting teaching based on evidence from assessment. However, disagreement with *assessment allows different students to get different instruction* suggests that this is not at a personal, student level, but rather that modification of teaching is carried out at the whole class level.

Factor 4, *assessment for school accountability*, is identical to that of G.T.L. Brown (2006), but responses to the items contributing to this factor did not necessarily show strong endorsement of this function of assessment (40% of participants did not agree that assessment is an accurate indicator of school quality and 45% did not agree that assessment is a good way to evaluate a school). This aligns with Wylie and Bonne (2016), who claimed secondary school teachers in New Zealand are reluctant to make this connection between assessment results and school quality. The range of responses to these items indicates that some participants did not consider assessment results a

useful measure of schools and agree with the argument that to use assessment in this way is damaging and unhelpful (Broadfoot, 2007; Madaus, Scriven & Stufflebeam, 1983; Nichols & Harris, 2016).

The factor structure and responses to the items show that participants hold both formative and summative conceptions of assessment and that formative use can be made of summative assessments. They are reluctant to agree that assessment information should be used to measure a school's quality, or that assessment is irrelevant.

Inter-correlations between the factors would support these claims. The positive correlation of .53 between *assessment awards qualifications* and *assessment improves learning* indicates that participants who endorse one of these conceptions also tend to endorse the other; they may consider that assessment can serve both formative and summative purposes. For NCEA internally-assessed standards this is certainly possible, as teachers decide on the nature of the internal assessment tasks, on the conditions for the assessment (in general), and on resubmission policies that will influence the formative opportunities available. For example, schools (or teachers) can allow students to resubmit assessments to improve grades which provides an opportunity for formative feedback to occur. Similarly, the type of task chosen may provide formative opportunities; for example, a portfolio of evidence gathered over time would allow for the summative function of assessment to be integrated with the formative function. Harlen (2006) and Maxwell (2004) speculated that it is possible to use internal assessment to fulfil both functions of assessment, and empirical studies from New Zealand (East, 2016) and Hong Kong (Yung, 2012) have reported that some teachers do provide multiple and ongoing opportunities throughout the year to attain the requirements of internal assessment. Teachers in this study who adopted this latter stance would consider that NCEA internal assessment can contribute formatively to student learning.

Factor 2 *assessment has weaknesses* has negative correlations with all of the three other factors. This is logical because, if participants believe assessment has useful functions such as awarding qualifications, improving learning, and indicating school quality, then it is obviously not irrelevant. These inter-correlations have similarities

to, and differences from, those previously reported. Research on conceptions of assessment undertaken with secondary school teachers internationally (Brown et al., 2011; Brown et al., 2009; Gebril & Brown, 2013) has similarly reported that using assessment results for accountability is a relevant use, as opposed to research undertaken with primary school teachers, which has shown that using assessment to measure students was more associated with assessment as being unfair (G.T.L. Brown, 2008).

Correlations with factor 4, *assessment for school accountability*, reflect the mixed agreements with items in this factor. On the one hand, there is a positive correlation of .53 with Factor 1, *assessment awards qualifications*, suggesting that participants believed that assessment results can be (or are) used as an indicator of school quality, in agreement with Egyptian secondary school teachers (Gebril & Brown, 2013). This is in contrast with their primary colleagues in New Zealand who rejected the notion that schools should be held accountable for their assessment results (G.T.L. Brown, 2011a) and with Wylie and Bonne (2016), who claimed that secondary teachers in New Zealand similarly rejected this function of assessment. On the other hand, the weak inter-correlation of .18 between *assessment results can indicate school quality* and *assessment improves learning* indicates that participants do not believe the more formative functions of assessment can be an indicator of school quality. There are conflicting pressures when teachers are making assessment judgements for student learning and the quality of their teaching and their school is also judged by those assessment results. Participants seemed to be in agreement with Broadfoot (2007) who signalled a fundamental incompatibility between these two functions of assessment. Making schools accountable for assessment results, Broadfoot claimed, is associated with performativity with an emphasis on teaching for better results rather than improving learning.

6.1.3 Participants' conceptions of assessment

According to analysis of the quantitative data, the group of teachers participating in this study displayed conceptions of assessment similar to, but not identical with, the more general sample of teachers in G.T.L. Brown's (2008, 2011a) research. In general, their conceptions of assessment align more closely with those of teachers in high-stakes qualifications jurisdictions (Brown et al., 2011; Brown et al., 2009; Gebril

& Brown, 2013). The differences from the primary teachers in G.T.L. Brown's (2008) and Brown and Lake's (2006) studies support Remesal's (2011) assertion that teachers' conceptions of assessment may be related to the ages of their students. The extraction of a different conception from G.T.L. Brown (2006), *assessment describes learning*, which describes both summative and formative beliefs about assessment, reinforces an ecological approach whereby teachers' conceptions of assessment are influenced by the assessment environment in which they work (Brown & Harris, 2009; Brown, Lake & Matters, 2011). Accounting and economics teachers teach almost exclusively at the senior level of secondary school and have a strong focus on assessing for qualifications, which includes carrying out a substantial amount of assessment for national qualifications. Despite this focus on qualifications, teachers in this study also showed a strong tendency to hold a conception of assessment that relates to the formative function of assessment; they agreed that internal assessment for qualifications could fulfil both formative and summative functions.

6.2 Discussion of qualitative themes

The qualitative aspect of this thesis explored reasons for teachers making the decisions they do about the internal assessment tasks they offer students. Analysis revealed five key descriptions of internal assessment: *assessment as pedagogy*; *assessment as motivator*; *assessment for educational outcomes*; *assessment as practice*; and *assessment is detrimental*.

6.2.1 Assessment as pedagogy

The theme *assessment as pedagogy* reflects a formative view of assessment – learning is a key purpose of assessment (Black & Wiliam, 1998; Harlen, 1994; Sadler, 1989). Some participants recognised the potential for internal assessment to be integrated with learning in authentic ways and strove to make it part of the teaching and learning process. As such, they were enacting the theoretical possibilities for the formative and summative functions of assessment to be integrated (Maxwell, 2004; Harlen, 2006). They described giving students ongoing feedback and scaffolding their learning to complete internal assessments. This process included teaching students how to self-assess against criteria, allowing them to collaborate with each other and giving them more than one opportunity to achieve. These actions exemplify several aspects of the

formative use of assessment: involving students in the process (Harlen, 2006); using feedback to improve the work (Taras, 2010); and identifying gaps in students' knowledge (Sadler, 1989). Actions were also indicative of a socio-cultural view of learning (Vygotsky, 1978) whereby external supports are a key element in learning and assessment, and teachers and students collaborate to obtain students' best performance (Gipps, 1994). Using portfolios and multiple opportunities to achieve standards can add validity to the teachers' judgements because they build a picture of student achievement over time (Davison & Leung, 2009; Harlen, 2004; Maxwell, 2004). While these actions may not meet the vision of Hipkins et al. (2016) that internal assessment within NCEA be enmeshed with learning, some participants in this study were integrating internal assessment with their pedagogy as much as they considered possible.

6.2.2 Assessment as motivator

Black and Wiliam (1998) emphasised the importance of involving students in the assessment process and Klenowski and Wyatt-Smith (2014) added to the socio-cultural view of assessment by including the proposition that assessment is used to engage students in their learning and to admit them to the community of learners. Thus, the theme *assessment as motivator* recognises the potential to connect assessment with students' interests which, in turn, would provide motivation for students to complete the tasks and achieve well (Hulleman, Durik, Schweigert, & Harackiewicz, 2008). Some participants understood the notion that interest in activities would lead to motivation (Deci & Ryan, 1985; Dewey, 1913) and that motivation and achievement are linked (Nicholls, 1979). Numerous examples were given of teachers contextualising assessment to students' local areas or interests, for example, through using local businesses for case studies and field trips, and drawing on standards from other subjects, such as Business Studies, to create interest and authentic contexts in which to study accounting and economics. These practices show that participants were enacting assessment policy (Ministry of Education, 2007, 2011) as intended; they were using the National Qualifications Framework to design and deliver programmes that attempt to "engage students and offer them appropriate learning pathways" (Ministry of Education, 2007, p. 41).

In addition, participants considered that achieving NCEA credits, and achieving them well, was a motivating factor for students. They ensured that students achieved some internally-assessed credits early in the year, recognising the relationship between success and motivation (Weiner, 1985). While performance approaches to achievement are linked to extrinsic motivation, there is evidence this approach can lead to more intrinsic motivation later in the learning (Meyer et al., 2007). Participants certainly observed that early success in NCEA led students to have more confidence and more willingness to try. They also believed that the ability to gain course and certificate endorsements was highly motivating for some students, as predicted by Meyer et al. (2007) and confirmed by findings of Meyer, McClure, Weir, Walkey, and McKenzie (2009) who reported that students considered endorsements important and found a “positive relationship of availability of endorsements to motivation” (p. 3).

In contrast to previous research (NZQA, 2014; PPTA, 2007) that raised concern about NCEA endorsements restricting flexible course design, participants in this study were increasingly innovative in their course design, and those who did not already use standards from other domains shared their plans to consider this in the future. Participants tended to offer more innovative and mixed courses at Levels 1 and 2 of NCEA, whereas at Level 3 a pure disciplinary course of either economics or accounting was the norm.

Harris and Brown (2009) also reported that teachers in their study talked about motivation being a purpose of assessment, but their participants conceived of this as external motivation whereby competition, pressure and praise were used to motivate students to work harder. In contrast, participants in this study seemed to be describing the desire to create intrinsic motivation; they did not mention competition nor pressure, but rather wanted to create a desire to complete and succeed by appealing to students’ interests, and by enhancing motivation through early success.

6.2.3 Assessment for educational outcomes

Assessment for educational outcomes reflects a summative view of assessment, in that assessment results provide opportunities for students in the future. This theme describes the rational, strategic choices that teachers make in order to maximise life chances for students. Key aims of NCEA were to achieve higher proportions of students staying in school and for more students to gain qualifications, particularly those who had previously been excluded (Hipkins et al., 2016; Hipkins & Spiller, 2012; Meyer et al., 2007). Participants certainly agreed that they must enable most, if not all, students to gain national qualifications with the highest possible grades, and they implemented many strategies to ensure this occurs. They were aware that students may perform better in internal assessment than in external examinations and increased internally-assessed opportunities while reducing the number of external standards students would attempt. Most participants advised their students to attempt fewer than the maximum number of three external examinations, citing there was no need for more credits or the difficulty of achieving high grades if all three were attempted. A further strategy included reducing the amount of assessment altogether by removing standards and topics that students find challenging, or uninteresting, or those the teacher deems unnecessary. The drive for fewer, high quality credits (i.e., those gained at Merit or Excellence level) was a result of the 2007 and 2010 changes to NCEA that allowed for certificate and course endorsements of Excellence and Merit (NZQA, n.d.d), and also of New Zealand universities requiring a minimum number of Merit and Excellence grades (in addition to University Entrance) for admission to restricted courses (Hipkins & Spiller, 2012).

Klenowski and Wyatt Smith (2014) described how teachers might adapt assessment to cater for students' diverse needs and to promote participation of students into the learning community. Some participants provided evidence that they attempted a more student-centred, socio-cultural approach to assessment by having differentiated assessment opportunities. They allowed students to choose additional or different standards against which they could be assessed, thus improving participation and outcomes for those students. They achieved this through providing extra tuition and blended learning techniques, which included using proprietary and teacher-made resources for students to work on semi-independently. These actions could also be

seen to approximate the vision of Hipkins et al. (2016) for NCEA which calls for schools and teachers to use the flexibility of NCEA to provide “lifeworthy learning for each and every student” (p. 206).

Rata and Taylor (2015) put forward a contrasting view with regard to offering different NCEA standards to different students, expressing concerns that restricting students’ access to epistemological knowledge is inequitable. They argued that offering differentiated knowledge to students who were considered to have lower ability was based on teachers’ value judgements and was likely to perpetuate social and economic inequities. Similarly, Riley (2014) hypothesised a correlation between school decile and the rigour of standards offered. Teachers in this study appeared to be aware of the need to balance the demands of meeting students’ needs while still teaching the discipline of the subject. Participants considered that students learn important life skills, such as financial literacy, from economics and accounting and consequently attempted to ensure that students achieved these. Some narrowing of the curriculum had occurred, in particular with many accounting teachers removing the accounting concepts topic, but teachers in both high and low decile schools had removed this topic; this decision was based on teachers’ perceived value of that topic and/or student interest, not necessarily on ideas about students’ ability to achieve the standard. The economics teachers tended to teach the whole curriculum at Levels 2 and 3 but at some point either they, or the students, decided which assessments students would attempt.

Participants who did teach the whole curriculum were attempting to ensure that students learned the epistemological knowledge of the disciplines. Rather than narrowing the curriculum to that which will be assessed, as described by Broadfoot (2007), Pedulla et al. (2003), and Stobart (2008), teachers in this study were concerned that students be provided with a strong pedagogic framing. They chose content to provide a sound grounding in the discipline (Bernstein, 2000, cited in Rata & Taylor, 2015). They were aware that students’ decisions can undermine this pedagogic aim and were concerned that when students chose which standards to attempt they could be denying themselves important disciplinary knowledge. As one teacher said, “I will choose not to do all the standards, then the students will choose not to do some, so in the end they haven’t done much economics at all”. Consequently, teachers attempted

to minimise student opt-out from assessment and adopted a range of practices to ensure maximum curriculum coverage.

Participants seemed aware of their students' possible next steps in life and did everything possible to enable their students to achieve these. If university was the chosen path, teachers ensured entrance requirements would be met, including those teaching in low decile schools – a finding that differs from previous research (Madjar, McKinley, Jensen, & Van Der Merwe, 2009; Turner, Irving, Li, & Yuan, 2010). Both Madjar et al. and Turner et al. reported that students from low decile schools did not achieve university entrance at the same rates as students from medium and high decile schools. Madjar et al. concluded that the flexibility and choice associated with NCEA led some students and their teachers to make inappropriate choices in terms of students' ability to gain University Entrance. There were only five participants from low decile schools, but all reported managing NCEA assessment to ensure their students met the necessary requirements. One gave an example of conducting internal assessments under test-like conditions so students could not opt out (as they did when they were expected to complete portfolios) and another always had additional internally-assessed standards “up my sleeve in case a Level 3 student needs it to go to university”. Nonetheless, this could be a case of espoused rather than actual practice.

Participants made a number of rational choices to ensure positive educational outcomes for their students. This included a differentiated and student-centred approach to structuring the curriculum and the subsequent NCEA assessment.

6.2.4 Assessment as practice and is detrimental

Assessment as practice reflects the major responsibility the participants have in implementing and managing the internally-assessed NCEA standards. Closely related to assessment practice are actions and beliefs, which reflect the detrimental or negative aspects of assessment; therefore, these two themes will be discussed concurrently. As pointed out earlier by NZQA (n.d.d), teachers have the freedom to create assessment tasks that suit their students and teaching programmes. The Ministry of Education and NZQA provide exemplar assessment tasks, assessment schedules, and examples of student work at the different NCEA levels of achievement, but there is no requirement to use these exemplar tasks. In addition, the Conditions of Assessment for internal

achievement standards (Ministry of Education, 2017) are liberal, suggesting that: flexibility can be exercised when gathering assessment evidence; separate assessment events are not needed for each standard; and the assessment method should not be the same as that for external examinations. Despite this freedom, most participants preferred to use or modify proprietary tasks, or the exemplars, rather than creating their own, a finding that aligns with previous literature reporting on teachers' NCEA internal assessment practices (Alison, 2005; Edwards, 2013; Kane & Fontaine, 2008; Moeed, 2010). Most reported using in-class tests for at least one internal assessment task. The key reasons given for this were perceived expectations from NZQA or their school management, their own professional confidence and safety, workload issues, and ensuring authenticity of students' work.

Professional confidence and safety

Negative experiences of external moderation and the subsequent actions taken by school management if moderation reports were considered to be 'poor' had eroded some participants' confidence and led to self-doubt; consequently, they worked to minimise professional risk by using, or strongly imitating, TKI exemplars. Many considered it too risky professionally to design original tasks only to have them "ripped apart by the moderator". These findings concur with previous research (Brooks, 2010; East, 2016; Mizutani, 2009; NZQA, 2014; Taylor et al., 2007; Wylie & Bonne, 2016) in which teachers have reported critical feedback from moderators, a process which Hipkins et al. (2016) suggested may make teachers risk-averse and reluctant to use innovative approaches. Dissonance between participants' ideas of acceptable assessment tasks and those of the moderator, whose judgements may be no more valid than those of the teacher (Johnson, 2013), had caused some participants to retreat to safer practice (Pedder & Opfer, 2013). Hipkins (2015) claims this practice is understandable, but unfortunate, because it is unlikely to achieve a variety of ways of gathering assessment evidence. Participants in this study reinforced the views of Hipkins et al. (2016) and Hipkins (2015) in that none shared what could be considered innovative ways of carrying out internal assessment, although some expressed a desire to do so, citing external demands to conform with moderators and school management as key barriers. Participants claimed that moderators actively discouraged innovative assessment design, instead recommending that teachers use, or adapt, the exemplars

provided by the Ministry of Education, stating that original tasks were unlikely to meet moderation requirements.

Research for NZQA itself (NZQA, 2014) and the Office of the Auditor General (2012) has also reported that moderation procedures have left teachers dispirited and feeling threatened when moderators do not agree with their judgements. Validity and reliability issues understandably have dominated the moderation process because internal assessment results contribute to national qualifications and so must be valid, and marking consistent among teachers, to ensure fairness for all students. However, participants perceived this process to be heavy-handed rather than educative, expressing feelings of resentment, anger and cynicism, rather than respect towards the moderators.

While some participants maligned the moderation processes, the subsequent actions taken by school management if moderation reports were deemed unacceptable was of as much concern. Participants expressed the opinion that their school management seemed to view not ‘passing’ moderation as an indication of professional incompetence and held teachers accountable. Despite the NZQA (2014) claim that the purpose of moderation is educative and intended to provide feedback to teachers and schools about assessment decisions, not just agreement rates, schools operationalise the feedback as ‘passing’ or ‘failing’ moderation. Moderation reports had ceased to be informative about ways in which teachers might improve their practice, and had become mechanisms by which teachers and schools exhibit accountability. All participants were held accountable if the moderator disagreed with their marking or assessments tasks. For example, Scotsaigh’s school “wanted to know what was going on”, Kirsty felt judged by moderation, and Mhairi had to justify herself to her school management even when she thought the moderator was wrong. The combined effect of seemingly negative feedback from moderation, and the associated negative reaction of school management, appear to be examples of what Codd (2005) described as a managerialist culture, whereby schools are more concerned with what can be recorded and documented about teaching and learning than with learning itself. According to Codd, managerialism “treats teachers as functionaries rather than professionals and thereby diminishes their autonomy and commitment to the values and principles of education” (p. 201). Some participants and their schools seemed to value ‘passing’

moderation more than providing pedagogically sound, meaningful and innovative internal assessment tasks, suggesting that the summative, accountability function of internal assessment was dominating any possible learning functions.

Hall (2005) had warned that although the NCEA assessment reform intended to involve teachers more in the assessment process, central control did not diminish; subsequent monitoring limited teachers' freedom. One could argue that central control has actually increased. Some participants who were confident to produce their own assessment materials did so in consultation with the moderator, and, rather than describing a collegial conversation, such consultation seemed more like advice-seeking with the power and control residing with the moderator. Further evidence of NZQA asserting power and control is in the provision of Best Practice Workshops. While the provision of these is welcomed by teachers (Hipkins, 2013; NZQA, 2014; Office of the Auditor General, 2012; PPTA; 2016), the fact that the national moderator is the facilitator of the workshops can further position him or her as the sole expert.

Professional support

In contrast to previous research on Best Practice Workshops which elicited generally favourable perceptions (Hipkins, 2013; NZQA, 2014; Office of the Auditor General, 2012; PPTA, 2016), participants in this study highlighted the inadequacy of professional support that is provided for implementing NCEA internal assessment. Some considered the Best Practice Workshops useful professional development, while others found them frustrating, describing situations in which the workshop descended into dissenting disarray. The useful aspects were meeting the national moderator, which had led to a more positive working relationship, and networking with other teachers. It seems that the purpose of achieving consistency was less successful with some questioning the ability of the moderator to provide sound professional advice and believing that changes in the national moderator led to expected standards changing. NZQA (2014) has reported similar concerns regarding "how the wisdom of the solo expert moderator gets transferred to the successor" (p. 12) and questioned the expertise that existing national moderators had with the 2010-2012 revised internal standards. Comments by participants, such as "Sometimes I think I have more experience than the moderator", "I wonder how they got the job", and "Just because

you have the name ‘moderator’, does that mean you know it all?’” certainly confirm previous opinions.

Workshop delivery ignores previous preferences expressed by teachers for a more personalised approach to NCEA professional development (Starkey et al., 2009; Taylor, Yates, Meyer, & Kinsella, 2011). The international literature (e.g., Dall’Alba & Sandberg, 2006; Hoekstra, Brekelmans, Beijgaard, & Korthagen, 2009) agrees that, because many different practices are occurring among teachers when reform is being embedded, there are many different professional development needs and therefore support for teacher learning needs to be differentiated. All participants agreed that they preferred the more individual approach of working with selected colleagues from other schools, citing networking with other colleagues as a valuable source of professional support.

Participants described moderation procedures and professional development for internal assessment as control rather than professional support, which resulted in feelings of self-doubt, anger, and distrust. Instead of teachers and the national moderator coming to shared understandings about how standards might be assessed, the national moderator has the power to determine this. This situation is less than ideal when the competency of the national moderator is called into question, and even if the national moderator is highly competent, participants questioned allowing this level of power to reside in one person. NZQA (2014) claims it has “well articulated processes for inducting new moderators and for supporting their practice” (p. 13) but also admits that “building teachers’ assessment practices relies on the concurrent building of moderator practices” (p. 13). An avenue for future research would be to provide empirical evidence regarding national moderators’ pedagogically sound (or otherwise) understandings of how internal assessment should or could be carried out and investigation into national moderators’ practice in relation to policy and standards.

Pragmatic decision-making

A further aspect of participants' internal assessment practice that seemed to be influenced by detrimental or negative aspects of assessment was the tendency for some, if not all, internal assessment tasks to be conducted as in-class tests. Despite the Conditions of Assessment (Ministry of Education, 2017) actively discouraging examination-like conditions for internal assessment, participants reported continued use of this method. This practice is the antithesis of the aims of internal assessment, which are to allow assessment to be integrated with teaching and learning and to assess a much greater range of skills (Davison & Leung, 2009; Harlen, 2006; Maxwell, 2004; NZQA, n.d.a). However, the practice is not unusual and aligns with findings on teachers' practices with internal assessment in Hong Kong (Yung, 2012) and other studies on the New Zealand system (East, 2016; Moeed, 2010). Yung claimed that teachers' beliefs seemed to be a key driver in how the teachers in his study operationalised internal assessment; however, participants in this study seemed to understand that alternative forms of assessment, such as portfolios, were more pedagogically defensible, but cited pragmatic reasons for not using them. East and Moeed similarly reported on the prevalence of pragmatics as a driver for decisions on how to assess the NCEA internally-assessed standards.

Pragmatic reasons that dominated participants' decisions to use tests for internal assessment were fear of student dishonesty and teacher workload issues. Student dishonesty, either through students sharing answers with each other, by accessing answers to the exemplars that are available on the NZQA website, or, in one school, students using online ghost-writing services, was cited as a key reason for conducting internal assessments as in-class tests. The increasing opportunities the internet affords for student dishonesty was a proposition put forward by several participants. Murdock, Stephens, and Grotewiel (2016) claim it is difficult to gauge the prevalence of academic cheating because of the reliance on institutions self-reporting, but there is general opinion that it is on the rise along with evidence that up to 50% of secondary school students in New Zealand admit to cheating or collaboration (Colmar Brunton, 2012, cited in Murdock et al.). NZQA (n.d.h) assigns the responsibility of ensuring the authenticity of students' work to assessors, who in this case are the teachers – a role they take seriously. Participants and their schools cited elaborate mechanisms to

ensure the authenticity of students' work, but for most, the use of an in-class test was seen as the most expedient way to ensure students' work was authentic.

However, not all participants expressed such strong opinions about student dishonesty. Those who did not, favoured more open methods of assessment that ensured authentic work through: knowing the students and their work well; providing on-going monitoring and feedback to students as they progressed; designing internal assessment tasks to appeal to student interests or allowing students to work on self-selected topics; and teaching students the ethics of plagiarism. Some authors have rationalised that a student's personality determines the level of dishonesty (Giluk & Postlethwaite, 2015), but others have suggested it could be a function of the classroom context with classrooms that value personal improvement and risk-taking, as opposed to competition and goal setting, displaying lower incidences of cheating (Anderman, Griesinger, & Westerfield, 1998; Murdock, Hale, & Weber, 2001).

Other researchers have claimed that students' attitudes towards cheating are affected by the quality of their relationships with teachers (Anderman, Freeman, & Mueller, 2007), and the pedagogical competence and personal commitment the teacher displays (Rabi, Patton, Fjortoft, & Zgarrick, 2006). Findings from this study would support these claims. Participants who reported carrying out all their internal assessment under test-like conditions also reported making minimal changes to the exemplars to prevent plagiarism (not to meet students' needs), allowed few or no resubmission opportunities, and some had a generally negative attitude towards NCEA. It is interesting that the seven participants who reported these actions and beliefs worked in high-decile schools with six of them in single-sex schools. These types of schools typically achieve more highly in NCEA (Thrupp & Alcorn, 2010) and could be seen to favour testing; in the 1990s they lobbied to maintain external examinations rather than internal assessment (Strachan, 2016, cited in Hipkins et al., 2016). This aligns with Anderman et al.'s (1998) and Murdock et al.'s (2001) propositions that the culture of schools plays a major role in fostering different rates of cheating.

Workload issues have been raised in many assessment jurisdictions that include an internal assessment component (e.g., East, 2016; Stanley et al., 2009; Yan, 2014; Wylie & Bonne, 2016; Yip & Cheung, 2005), and participants in this study were no

exception. Managing internal assessment creates considerable workload as teachers must write or adapt multiple assessment tasks, which they subsequently mark. Therefore, workload considerations influence the choice of internal assessment tasks. The use of in-class tests, in conjunction with other forms of assessment (e.g., portfolios), was justified by some participants on the basis that the combination of assessment types balances students' opportunities to achieve with their own workloads. Xu and Brown (2016) discuss similar compromises required to ensure that teachers' assessment practice is feasible. Several participants agreed with the example provided in Xu and Brown that large class sizes limit the manageability of too much portfolio-based assessment and limit their provision of resubmission opportunities.

A further concern is that excessive workload can lead to negative feelings about internal assessment (Jaba, 2013; Majid, 2011). Some participants expressed such negative feelings and a desire to revert to more external assessment, but said they were prevented from doing so by school policies. The 2010-2012 changes to NCEA standards, which increased the internal assessment component to at least 50% for each subject, had led to resentment towards NZQA. For some participants, this change was regarded as a cost-saving measure on the part of NZQA; by reducing the external component, NZQA would pay fewer external markers, a cost that some saw was transferred to teachers who now had more marking responsibility but no corresponding monetary reward. Whatever the validity of this claim, it points to negative attitudes and distrust that have built towards NZQA. Additionally, when teachers perceive their workload to be excessive they may not produce quality work (Qian, 2014), and actions such as participants making minimal, or no, changes to exemplars, and conducting internal assessment as tests are indicative of this occurring.

Performance expectations

The increasing pressure from schools for teachers to meet performance targets was another negative impact on participants' practice. Participants and their colleagues were held accountable by the school's management for their students' achievement rates in NCEA. As one rather bluntly said, "you get knifed in the back by your own school for having Merits, not Excellences" and another suggested his predecessor had been "moved on" due to students not meeting expected achievement rates. Some participants reported that their schools had expectations that certain percentages of students would achieve

Merit and Excellence at each NCEA level and that their school's achievement would at least be comparable with other similar schools. Hipkins et al. (2016) claimed that teachers were indicating that schools set them individual achievement rates and participants in this research have supported this. An issue that can arise from having performance expectations is an inclination to teach to the test (Wyse & Torrance, 2009), which some participants reported trying to avoid, whereas others admitted that at times they felt they were "teaching the students to be assessed". It seems that, as Broadfoot (2007) warned, if assessment data are used to evaluate schools and teachers, a culture of performativity is likely to arise.

Achieving certain pass rates could be considered a legitimate aim of schools, as teachers should have accountability for their practice, but if the consequences are too high this could lead to beliefs that this function of assessment is irrelevant (G.T.L. Brown, 2008). Participants were not in favour of this practice and felt under pressure to meet these demands. The literature has reported on the stress students experience around assessment and performing well (e.g. Locker & Cropley, 2004), but less is written about the stress teachers experience when preparing students for assessment. In addition to workload stress associated with NCEA, participants also experienced stress with respect to meeting achievement targets.

6.2.5 Internal assessment practices

The preceding discussion has revealed that teachers' internal assessment practices are multiple and varied. Their practices can be conceptualised on a continuum from complying with an assessment culture to enacting a testing culture (Wolf, Bixby, Glenn, & Gardner, 1991). Internal assessment within a testing culture tended to be in-class, separate assessment events with little or no feedback or reassessment opportunities. In an assessment culture, portfolios with ongoing feedback and reassessment opportunities were provided, practices that align more closely to the philosophy behind incorporating internal assessment into qualifications (Davison & Leung, 2009; Harlen, 2006; Maxwell, 2004). Although, even within an assessment culture, the NCEA internal assessment was separated from normal classroom learning. East (2016) reported there was a perception among teachers that high-stakes assessment needs to be separated from normal work and claimed teachers need support to accept the validity of lesson-embedded evidence. Data from this research support that claim but suggest that the

national moderators also need such guidance. Their advice that teachers use internal assessment tasks that closely resemble the exemplars reinforces this interpretation. If national moderators tell teachers that teacher-created tasks are unlikely to ‘pass’ moderation, and teachers fear for their professional safety if moderation is not passed, then a cycle of safe, rather than innovative, practice is encouraged.

Participants did tend to situate themselves in either an assessment or a testing paradigm, but these behaviours have been described as a continuum; not all individual teacher actions were typified by the extremes. Most of the participants ($n=15$) described behaviours that were more at an assessment culture end of the continuum, behaviour that has been analogised as acting as ushers because they opened doors to create opportunities for students to achieve. At the other extreme, seven participants behaved more like gatekeepers in that they treated internal assessment as extensions of examinations. However, participants were not necessarily dichotomous within the continuum, illustrating the complexity of behaviours and beliefs associated with assessment (Bonner, 2016; Brown, 2011a; Harris & Brown, 2009; Remesal, 2011; Wyatt-Smith, Klenowski, & Gunn, 2010). The aim, however, was not an attempt to pigeon-hole the participants into a narrow assessment identity (Looney et al., 2017) but to focus on assessment behaviours that clustered around certain ideas and behaviours. As with previous literature (e.g. Adie, 2013; Davison, 2004; Horsley, 2012; Shohamy, 2001) metaphors were used to describe these behaviours.

East (2016), Mizutani (2009) and Moeed (2010) have described varying practices of teachers in relation to NCEA, and the findings from this study concur that different teachers actualise internal assessment in different ways. The findings from this study also show teachers themselves vary in their assessment practices. At times, individual teachers would capitalise on the opportunities afforded by internal assessment and provide formative and innovative opportunities. On other occasions, and with different standards, they would not.

CHAPTER 7

Discussion – Converging the data strands

The confirmatory factor analysis showed an inadmissible fit of the data to G.T.L. Brown's (2006) model, although the confirmatory and exploratory factor analysis reproduced aspects of the model. When the quantitative and qualitative data are considered together, greater similarities with G.T.L. Brown's model emerge. Convergence of the data from both strands of the research shows that participants in this study agree with most of the conceptions of assessment as identified by G.T.L. Brown. On the other hand, one conception was not reproduced and a different conception of assessment related to being an assessor for qualifications emerged. The interviews provide further explanations for patterns and relationships extracted from the qualitative analysis.

Table 7.1 summarises the quantitative factors and qualitative themes and how they converge and complement each other

Table 7.1

Convergence and complements among quantitative factors and qualitative themes

Quantitative factors	Overarching conceptions	Qualitative themes
<p>Assessment for improvement Assessment provides feedback to students about their performance, teachers modify teaching to meet needs identified from assessment and assessment can be integrated with teaching practice.</p> <p>Assessment describes learning Assessment establishes what students have learned, is not ignored, nor unfair, measures higher order thinking skills and determines if students meet qualifications.</p> <p>Assessment has weaknesses Assessment forces teachers to teach in ways against their beliefs, interferes with teaching, and is imprecise.</p> <p>Assessment for school accountability Assessment can be an indicator of school quality.</p>	<p>Assessment is for learning Assessment for improvement. Assessment as pedagogy. Assessment as motivator.</p> <p>Assessment is for qualifications Assessment for educational outcomes. Assessment describes learning.</p> <p>Assessment is detrimental Assessment has weaknesses. Assessment is detrimental.</p> <p>Assessment is for accountability Assessment as practice. Assessment for school accountability.</p>	<p>Assessment as pedagogy Assessment is part of the teaching and learning process, but achieving integration of NCEA into learning is not straightforward.</p> <p>Assessment as a motivator Internal assessment can be used to motivate students and to connect with their outside interests.</p> <p>Assessment for educational outcomes Teachers work so that students achieve NCEA with maximum grades, show mastery of subjects and are prepared for university etc.</p> <p>Assessment is detrimental Aspects of assessment are irrelevant and not useful to teaching and learning and/or damaging to working conditions and professionalism</p> <p>Assessment as practice Assessment is part of teaching practice, which is influenced by professional development.</p>

7.1 Overarching conceptions

7.1.1 *Assessment is for learning*

The qualitative and quantitative data showed that teachers participating in this study hold a strong agreement with the formative, or improvement, conception of assessment. The qualitative themes *assessment as pedagogy* and *assessment as motivator* align with the quantitative factor *assessment for improvement* to support an overall belief that assessment is used to improve student learning. Items that contributed to the factor *assessment is for improvement* – which are that *assessment feeds back to students their learning needs, information modifies ongoing teaching of students, is integrated with teaching, provides feedback about their performance, and allows different students to get different instruction* – were reinforced in the qualitative interviews, whereby teachers described a range of strategies such as giving students feedback, new work and extra tuition to their improve learning. These actions align with previous literature on the formative use of assessment. For example, feeding back to students their learning needs and using assessment information to modify teaching are considered prerequisites for the formative use of assessment (Black, Harrison et al., 2002; Sadler, 1989). This strong association with an improvement conception of assessment aligns with previous research on teachers' conceptions of assessment (G.T.L. Brown, 2002, 2004; Brown & Gao, 2015; Brown et al., 2011; Brown et al., 2009; Brown & Michaelides, 2011; Chen & Brown, 2013; Remesal, 2011), and reflects New Zealand's formative assessment culture (Ministry of Education, 2007, 2011).

Despite the belief that assessment should improve students' learning, and all participants being able to explain the formative function of assessment, findings revealed that some participants had difficulty conceptualising how to implement the formative function of assessment in the context of internal assessment for NCEA. The quantitative data showed conflict, some participants generally disagreeing that assessment interferes with teaching and agreeing that it is integrated with teaching, whereas others did not. The qualitative interviews can provide some explanation for this contrary finding. Participants experienced tensions with regard to how much (if any) feedback they were able to give students, and described varying practices with regard to the types of internally-assessed tasks they provided for the students. As a result, at times some teachers were giving students feedback on internally-

assessed standards and incorporating these into classroom learning, but on other occasions they were not. Other participants rarely gave feedback and tended to assess internally-assessed standards through tests. These different practices can account for the mixed beliefs that feedback on internal assessment may, or may not, provide opportunities for student learning. Black and Wiliam (1998) and Broadfoot (2007), albeit not referring to NCEA, predicted that teachers would experience tensions in implementing both the formative and the summative functions of assessment, and this was the case for participants in this study. Bonner (2016) agrees that teachers experience ongoing tension and conflict trying to balance the expectations of external assessment with “beliefs that support AfL” (p.35).

It is not surprising that tensions and confusion existed among the participants; NZQA itself gives contradictory advice on the administration of internal assessment. Information from NZQA (n.d.i) supports a belief that teachers can provide feedback over a period of time as students work on portfolios or performance tasks, but suggests that the feedback should “become less specific the closer the student is to the submission date”. In terms of giving students resubmission opportunities, NZQA categorically states that any resubmission “must take place before the teacher gives any feedback to the whole class (or any student)”; however, further assessment opportunities can be preceded by more teaching if a new assessment opportunity is provided. This advice is clearly supporting the use of a separate internal assessment task that, at some point, will be submitted for marking and it is also actively discouraging using internal assessment in a formative way by not allowing feedback to be given. However, this is contrary to information provided on other sections of the NZQA website (e.g., NZQA, n.d.j) where comments such “Learning is not increased by repeated summative assessment, but by extensive feed forward and feedback” and “where evidence is accumulated over time, assessors will provide feedback at regular intervals during the preparation for a final version for assessment”. Teachers referring to these different web pages would be left confused as to acceptable practice for internal assessment and it appears there is little support or guidance on how teachers might integrate assessing internally-assessed standards more with teaching. Participants themselves had to work out how, and if, they could integrate assessment with learning, but innovative and experimental internal assessment design was considered professionally risky.

7.1.2 *Assessment is for qualifications*

The exploratory factor analysis identified a dimension – *assessment describes/measures learning* – that appears to reflect the significant role participants play as assessors for NCEA, with which the theme *assessment is for educational outcomes* converges. The *assessment describes/measures learning* factor included the items *assessment determines if students meet qualifications standards*, *assessment is a way to determine how much students have learned*, *assessment results are trustworthy*, *assessment establishes what students have learned*, *assessment measures students' higher order thinking skills*, *assessment results are consistent*, *assessment results can be depended on*, *assessment helps students improve their learning*, *assessment results are not ignored nor unfair*. All of these could relate to teachers implementing internal assessment for NCEA, and the theme *assessment is for educational outcomes* refers to strategies teachers adopted to enable students to gain qualifications with the highest possible grades, as well as to prepare them for their future lives. Together these ideas signal the role that teachers represented in this study have in awarding, and teaching students to achieve, national qualifications.

This conception of assessment has not been previously reported in the literature, but it is in some respects similar to G.T.L. Brown's (2006) student accountability conception in that it includes the item *assessment determines if students meet qualifications standards*, and Remesal (2011) reported a "societal-accreditation" conception of assessment. G.T.L. Brown's (2008) *student accountability* conception holds "students individually accountable for their learning" (p. 23) whereas the conception derived from this research also incorporates teachers' accountability towards that learning. Remesal's "societal-accreditation" conception also has elements of student accountability and summative functions of assessment, but a fundamental difference is that the conception derived from this study includes the teachers' role in awarding qualifications and encompasses the idea that assessment for qualifications can be integrated with classroom learning. Remesal claimed that her "extreme pedagogical" and "societal-accreditation" conceptions were two distinct aspects of teachers' conceptions, whereas the *assessment is for qualifications* conception acknowledges that the summative purpose of assessment can help students learn, and that through assessment teachers manage students' learning, so that students achieve qualifications.

This conception also bears similarities with previous studies (e.g. Brown et al., 2009; Brown et al., 2011; Hui, Brown & Chan, 2017) where teachers had conceptions of assessment in which improvement is strongly linked to making students accountable. A further similarity is with Brown, Chaudhry & Dhamija (2015) who found that accountability was a driving factor in teacher decision-making with regards assessment.

This conception also reflects a summative purpose of assessment, recognises that this purpose of assessment is an integral part of the New Zealand secondary education system, and that it has positive outcomes for students in the form of qualifications attainment (Madaus & Russell, 2011). Participants acknowledged the importance to students of gaining such qualifications and the profound consequences of not achieving these (Harlen, 2006; Sadler, 1989). The co-dimensionality of the item *assessment helps students improve their learning*, along with assessment strategies participants discussed in the interviews such as using portfolios and/or allowing students to resubmit assessments, show that internal assessment for summative purposes can be integrated with teaching and learning as Davison and Leung (2009) claimed.

7.1.3 Assessment is detrimental

The quantitative data reproduced G.T.L. Brown's (2006) *assessment is irrelevant* conception but there was low agreement with the individual items relating to that factor; which are *that assessment forces teachers to teach against their beliefs, interferes with teaching, is imprecise, treated cautiously, and made little use of*. Negative views towards assessment expressed in the interviews, such as there being too much assessment, assessment considered to be testing, and the teachers' vulnerability with regard to internal assessment do not align directly with G.T.L. Brown's irrelevant conception, but show some similarities. For example, too much assessment, and assessment interfering with teaching, are similar concepts, and echo the thought that aspects of summative assessment are not useful to learning and teaching, supporting the claim of incompatibility between the formative and summative functions of assessment (Black & Wiliam, 1998; Sadler, 1989).

Participants' vulnerability regarding internally-assessed standards is more reminiscent of Stiggins' (1995) fear of assessment that results from accumulated negative feelings towards assessment. These negative feelings have their origins in teachers' experience of assessment as students, and can carry on into their professional lives if they have to prepare students for assessments they may not agree with, or have to assess in ways they do not believe in.

Moderation processes for NCEA have certainly led to increased negative feelings about internal assessment, and rather than the process being educative for teachers, it has often resulted in resentment if there is disagreement from the moderator. According to Stiggins (1995), fear of assessment could lead to barriers that prevent teachers from analysing and improving their assessment practice. Therefore, participants who resented, rather than learned from, disagreement could be setting up barriers to improved practice. The impact of imposing accountability mechanisms on top of assessing for NCEA is having a detrimental impact on teachers' beliefs about assessment.

7.1.4 Assessment is for accountability

The quantitative data reproduced G.T.L. Brown's (2006) factor that *assessment is for school accountability* and, while the theme *assessment as practice* does not converge exactly with the quantitative factor, aspects of the factor and the theme do complement each other. Participants described actions of their schools indicating that assessment results are important indicators of accountability. Mutch (2012) has similarly reported that despite New Zealand not having official league tables, school management recognises that assessment results play a major role in stakeholders' judgements of the quality of schools. Mutch describes a chain of accountability from taxpayers to the government, government to the Ministry of Education, Ministry of Education to schools and from schools to teachers, which aligns with aspects of accountability described by participants. Schools have expectations of certain percentages of students passing various levels of NCEA and expectations of Excellence and Merit grades. Schools also expect teachers to write valid internal assessment tasks, mark students' work accurately according to the national standard (as determined by the moderators), and hold teachers accountable for 'poor' moderation reports. Qualitative data can explain the varied responses on the surveys to the school accountability items; participants recognised that

schools put assessment data to accountability uses, but did not necessarily agree that this is appropriate. For example, Failbhe's comment about pressure from the school for students to achieve Excellence and not Merit grades, Tearlag describing the pressure to achieve good grades so as to attract new students to the school, and Barra commenting that his teaching could be called into question if students achieved poor results.

Interestingly, neither data source revealed a *student accountability* conception of assessment and the Heywood-case correlation (i.e., greater than 1) between the *student accountability* factor and the *improvement* factor contributed to G.T.L. Brown's (2006) model being rejected as a good fit to the present data. This differs from previous research in New Zealand and Australia in which this conception has been reproduced (G.T.L. Brown, 2004, 2006, 2011; Brown, Lake, & Matters, 2011), but is consistent with studies of international high-stakes assessment regimes (Brown et al., 2009; Gebril & Brown, 2013), and also with a smaller qualitative study from New Zealand (Harris & Brown, 2009). Two items that G.T.L. Brown (2006) considered indicators of student accountability (*assessment places students in categories*, and *assessment is assigning a grade or level to students*) did not load on any factors in the exploratory factor analysis, and *assessment determines if students meet qualifications standards* loaded to an *assessment describes learning* factor. The lack of association among these items in the EFA suggests they were not measuring the same construct in the participants' minds.

In the interviews, participants did not give a sense of expecting students to be individually responsible for their own learning; on the contrary, they described elaborate mechanisms that they instituted to ensure that students achieved NCEA credits. Many indicated that students needed guidance and surveillance to complete internally-assessed standards and, if left to their own devices, many would not submit and therefore not achieve. As with Harris and Brown (2009), participants in this study described their practices in terms of their roles in improving student learning and achievement, rather than expecting students to be accountable. This finding differs from G.T.L. Brown's (2008) claim that an indicator of student accountability is the increasing role teachers give to students to determine their own learning. Participants in this study gave little indication that they allowed students substantial control; rather, they tended to decide which standards to assess and design their programmes to minimise what they called

“opt-out”. These actions are understandable if schools have expectations of individual teacher accountability because it is in the teachers’ professional interests to ensure high achievement rates for their students. Many of their actions with regards NCEA internal assessment seem to indicate that teachers are ensuring their work achieves the expected outcomes as indicated by school management and government policy. This aligns strongly with the power of external accountability as discussed by Lerner and Tetlock (1999) who claimed that accountability is a complex construct that produces a range of effects – only some of which are beneficial.

It is possible that by not expecting students themselves to be responsible that this stance is a further indication of the formative assessment culture in New Zealand schools. According to Black et al. (2002) formative assessment require teachers to use assessment information to alter their teaching to improve student learning, rather than expecting students to be entirely responsible.

Convergence of the data from both strands of the research has shown similarities with, and differences from, previous research on teachers’ conceptions of assessment. The differences reinforce the findings of Brown and Harris (2009), Brown, Lake, and Matters (2011) and Remesal (2011) who have argued that teachers’ conceptions of assessment are consistent with their use of assessment. In particular, the conceptions of assessment for participants in this study reflect the substantial role they have in assessing students for national qualifications.

7.2 Influences on teachers’ internal assessment practice

Participants revealed a range of influences on their internal assessment practice, as well as tensions that can arise when teachers are responsible for awarding qualifications. Frameworks have been proposed to understand teachers’ assessment practice more clearly. Fulmer, Lee, and Tan (2015) conjectured that there are three levels of influence on teachers’ assessment beliefs and practices: the micro-, meso- and macro-levels. The macro-level focuses on national policies and cultural influences, the meso-level on factors related to the school and the school community with the micro-level residing with individual teachers’ characteristics. Xu and Brown (2016) provided a framework whereby teachers’ knowledge of assessment provides a basis for their actions and is

filtered by their conceptions of assessment. According to Xu and Brown, only knowledge that is compatible with conceptions is accepted and used, and assessment practice is also influenced by micro- and macro-contexts which results in teachers having to make compromises that lead to a teacher assessment identity. Xu and Brown's framework is cyclical, so a teacher's assessment identity will, in turn, influence the compromises they make. Looney et al. (2017) extend these frameworks to include how teachers feel about assessment, that is, do they have the confidence to apply their knowledge and skills of assessment in the ways they prefer? Alternatively, they have skills, knowledge and confidence but may not believe certain assessment processes are effective. Looney et al. claim teachers' self-efficacy and their sense of control, in terms of assessment, are determinates of their assessment practice. The internal assessment practices of participants and their conceptions of assessment align with these frameworks. Their knowledge, skills, beliefs and feelings about assessment were reflected in the decisions they made in response to accountability pressures imposed by the micro-, meso- and macro-contexts within which they worked, and differences in these appear to result in different assessment practices.

7.2.1 The macro-context

Participants operate within the macro-context of the New Zealand education system and, within that, a macro-subsystem of national school level qualifications. Two overarching conceptions reflect this macro-context: *assessment is for learning* aligns with the formative assessment culture of the New Zealand education system, and *assessment is for qualifications* is indicative of the considerable role participants have in assessing for NCEA. The factor and thematic analysis found common responses to the macro-system but, in addition, the interviews revealed that teachers react individually to the influences of the macro-system. Teachers are compelled to comply with changes suggested by the national moderators, but attitudes towards the required changes varied. Some complied willingly, possibly acknowledging errors on their part, whereas others believed that the moderator is often wrong, but had no choice but to make the indicated changes.

From the participants' points of view, the macro-system of NCEA (including moderation) determined whether they were considered competent or incompetent assessors, which was framed as 'passing moderation' (or not). This feedback is public; school managers and NZQA personnel have access to moderation reports, which can

increase feelings of vulnerability, and as such has an impact on teachers' assessment identities. These common and individual responses to the macro-system support Xu and Brown's (2016) theorisation that teachers make assessment compromises that vary by teacher and context, but also there can be generalised decisions that are applicable to a wider range of teachers.

7.2.2 *The meso-level context*

According to Fulmer et al. (2015), the meso-level is the school community in which the teachers work and asserted that in the field of assessment research "relatively little prior work has paid explicit attention to these influences" (p. 485). Two of the conceptions derived from this research reflect the influence of the meso-system on those conceptions. The conception *assessment is detrimental* represents negative accountability influences from the macro-context, but also the strong impact of accountability expectations within the meso-level of the school. *Assessment is for accountability* embodies the culture of competition among New Zealand schools that has emerged since 1989. The reforms to New Zealand schools that began in 1989, known as Tomorrow's Schools, focussed on granting schools more autonomy, but also increased their accountability. More emphasis was put on competition among schools for students, with funding based on per-student formulas (Wylie, 1999). Schools assign this accountability to individual teachers in their expectations of NCEA performance.

The meso-level of the school imposed expectations with regard to the implementation of NCEA internal assessment, for example, there were policies on how the authenticity of student work should be guaranteed and the amount of internal assessment that could, or could not, be offered. Accountability towards school management was a strong driver influencing the choice of assessment tasks and the conditions under which these were completed. Practices within schools were shared and adopted by teachers even if they were not policy. For example, Scotaigh knew that his colleagues included part of the internal assessment in the mid-year examinations and indicated he would adopt that practice the following year. Participants who work in schools with departments of several economics or accounting teachers reported being required to comply with the demands of the department, for example, all students at a year level having to attempt the same internal task, at the same time, regardless of the suitability of the task or their readiness for it. A general pattern emerged with regard to the impact of the school type

on participants' assessment identity: those who aligned more with a testing regime tended to work in single-sex, high-decile schools, whereas those whose assessment identity was more associated with an assessment paradigm (Wolf et al., 1991) were more likely to be teaching in a mid to low-decile, state co-educational school.

Brown and Harris (2009) have similarly identified differing impacts of meso-level differences in the use of assessment data in primary schools and Remesal (2011) speculated that the school level of the students influences teachers' conceptions of assessment. However, the bulk of previous literature on conceptions of assessment has emphasised the impact of macro-level national assessment policy and culture on teachers' conceptions of assessment (Brown et al., 2015; Brown & Gao, 2015; Brown et al., 2011; Brown et al., 2009; Brown & Michaelides, 2011; Chen & Brown, 2013; Gebril & Brown, 2013). For example, Gebril and Brown attributed differences in Egyptian teachers' conceptions of assessment from New Zealand teachers to the high-stakes examination system in Egypt, and similarly in China, Brown et al. claimed Chinese cultural values and practices were responsible for differences from Western teachers. However, within a culture, teachers are not a homogeneous group and this research has shown that the meso-context of individual schools has an important influence on teachers' conceptions of assessment and their NCEA internal assessment practices.

In New Zealand it may not be possible to classify teachers as one cultural group; even among the 21 teachers interviewed for this research two identified as Indian, one as English and one as Scottish. Also, since 1990, New Zealand secondary schools have been self-governing, and have been encouraged to compete against each other for students (Wylie, 2012). As a result, schools have marketed themselves as having a distinct culture, for example non-uniformed, co-educational schools who portray a modern outlook, compared with uniformed, single-sex schools who espouse to values that are more traditional. Teachers within different secondary schools may adopt the beliefs and values of that school (or alternatively work in schools that are congruent with their own beliefs) rather than adopting more generic system-wide values. This research has shown that the meso-level of the school seems to be the over-riding influence on teachers' internal assessment practice.

The macro-system of NCEA and the meso-level of school actions, policy and accountability measures were key determinants of participants' feelings (confidence) and their perceived levels of skill (knowledge). These actions either affirmed teachers' self-efficacy as assessors or raised doubts and reduced confidence. Xu and Brown (2016) conceptualised teachers' assessment practice as a constant compromise between their knowledge and beliefs about assessment, and the contexts within which they work. Participants described many compromises among the competing influences on their assessment practice that support this claim. They also described tensions in making these compromises. The macro-context of the New Zealand education system advocates the use of assessment to improve teaching and learning, yet a sub-set of that system, NCEA, measures achievement and awards qualifications. Furthermore, an aspect of the macro-system uses NCEA pass rates to measure the effectiveness of education in New Zealand (New Zealand Government, 2012) and NCEA results are used as an accountability tool. Teachers in this study work in these contexts and experience the tensions between using assessment for formative purposes and at the same time for summative qualifications. These opposing purposes of assessment influence their conceptions of assessment and their assessment practices. Additionally, accountability measures within this macro-system, and the meso-system of their school, affected their knowledge of, and feelings and beliefs about themselves as assessors. Knowledge, beliefs, feelings and contexts differed between participants; therefore, the compromises reached varied by context and by teacher (Xu & Brown, 2016).

7.2.3 Beliefs about assessment

Participants revealed that their actual practices did not always align with their preferred practices and that external influences, such as NZQA requirements (or perceived requirements), authenticity of student work, school policy and workload dominated the rationales for their actions. Literature on teachers' beliefs (e.g., Bandura, 1997; Fives & Buehl, 2016) and teachers' conceptions of assessment (G.T.L. Brown, 2002) contends that beliefs and conceptions guide teachers' actions. Participants in this study revealed that this is not always the case and that external accountability measures can over-ride beliefs in terms of influencing the practices teachers adopt. Rokeach (1968) explains that newly-acquired beliefs are less likely to influence practice and it may be that for some teachers, beliefs relating to internal assessment are relatively new, rather than long-held. Sannino (2008) describes these as dominant and non-dominant

teaching behaviours. Dominant behaviours are those that are tried-and-true and well supported, whereas non-dominant behaviours relate to new initiatives that require change to existing practices and may not be adequately supported. While internal assessment for NCEA can no longer be considered a new initiative, it was a change to existing practice for most participants, and beliefs as to how it should be carried out are relatively newly acquired. Fives and Buehl (2016) explain how explicit and implicit beliefs either knowingly, or unknowingly, affect teachers' decisions. It is possible that participants' implicit beliefs determine their acceptance of, and compliance with, external accountability influences and that such beliefs explain some of the variation in individual participants' practices. It must be considered that teachers do not always have the autonomy to act according to their beliefs; in the present context teachers must balance their beliefs with the systemic realities of NCEA requirements.

This study has shown that there is not a straightforward relationship between teachers' beliefs and their internal assessment practices, because beliefs are only one influence on that practice. The macro- and meso-systems within which teachers operate are also dynamic and unstable; NCEA itself is not static. Numerous changes have occurred, such as the introduction of endorsements, the change in the relative proportions of internal and external assessments, changes in the nature and number of credits for standards, and changes to moderation procedures and personnel. The number of changing and conflicting influences in play at any point in time have resulted in a wide range of practices between teachers, and for individual teachers over time.

7.3 Enacting and embedding assessment reform

Educational change does not always occur as predicted because, as Senge (1990) claimed, the success of educational reform hinges on what individual teachers do in the classroom. The present research has shown that teachers' practices vary, and that teachers themselves are not consistent in their application of change. At its launch, NZQA (n.d.a) claimed that the internal assessment component of NCEA would provide a more comprehensive record of students' achievement, and give students credit for a greater range of skills and knowledge. For this to occur teachers needed to significantly change their practice and thinking about assessment (Hipkins et al.,

2016); however, this study has shown that teachers have many unstable reactions to change and a simple, linear, cause and effect explanation is not possible. Some participants described changes they have made and continue to make to their assessment practices that allow the theoretical possibilities of internal assessment to be realised, and students are given credit for a much greater range of knowledge and skills than they were under the previous assessment system. But, even for individual teachers, allowing for the theoretical possibilities of internal assessment was not a consistent practice; at times teachers limited possibilities for students by carrying out internal assessment as if they were external examinations.

A complex picture of teachers' assessment practices and the reasons for them emerged. These mixed reactions reflect the complexity of assessing students' learning, while also awarding national qualifications. They also reflect the difficulty of enacting and embedding educational reform and support Fullan's (2001) claim that the multi-dimensional nature of change accounts for mixed outcomes of educational reform. Fullan argued that there are at least three dimensions involved in the implementation of new policies and that teachers may alter none, one, two or all three dimensions. In the context of assessment reform these would be new assessment materials or technologies, new teaching and assessment strategies, and an alteration of beliefs about assessment. Hoekstra et al. (2009) added that during reform some teachers successfully integrate change but others actively resist it. The practices of teachers in this study show agreement with Fullan and Hoekstra et al., but add that making change is not as simple as having the motivation, beliefs and knowledge to do so. Teachers also need permission from the macro- and meso-systems within which they work and they need professional support to make such changes.

CHAPTER 8

Conclusion

8.1 Summary of the study

This study used an exploratory, sequential mixed methods research design to investigate the relationships between economics and accounting teachers' beliefs about assessment and their practices in relation to NCEA internally-assessed achievement standards. Participants' conceptions of assessment were explored through the TCoAIIIA Inventory (G.T.L. Brown, 2006). The quantitative data gathered by this inventory were analysed using confirmatory and exploratory factor analysis. Interviews were carried out with 21 volunteers to investigate further their practices for the internally-assessed component of NCEA and the reasons for making the decisions they do about the internal assessment tasks they offer students. The qualitative interviews were analysed thematically. The quantitative factors and the qualitative themes converged to describe the overarching conceptions of assessment of the participants. Interview data were also used to explain findings from the quantitative strand of the research.

8.2 Research outcomes

8.2.1 Conceptions of assessment

Participants revealed four overarching conceptions of assessment: *assessment is for learning*; *assessment is for qualifications*; *assessment is for accountability*; and *assessment is detrimental*. These conceptions have similarities with, and differences from, existing findings from research on teachers' conceptions of assessment.

The conceptions *assessment is for learning* and *assessment is detrimental* align strongly with G.T.L. Brown's (2006) conceptions of *assessment for improvement of teaching and learning* and *assessment is irrelevant*. These conceptions have also been found internationally (Brown et al., 2015; Brown & Gao, 2015; Brown et al., 2011; Brown et al., 2009; Brown & Michaelides, 2011; Chen & Brown, 2013; Gebril & Brown, 2013) which confirms that teachers recognise the importance of the formative

use of assessment, but also believe assessment has irrelevant aspects. While it might seem contradictory that teachers hold these conceptions simultaneously, it is rational for teachers to think in this way because there are multiple, and sometimes competing, purposes associated with assessment. As discussed by Brown and Michaelides (2011), teachers can think assessment is essential to teaching and learning, but at the same time be aware of the potential for the misuse of assessment results.

The *assessment is for accountability* conception reflects the idea that schools and teachers are held accountable for assessment results, which has some similarities with previous literature on conceptions of assessment (Brown et al., 2009; Gebril & Brown, 2013; Harris & Brown, 2009). A key difference is that the accountability conception found in this research did not include ideas of student accountability. This finding is consistent with an assessment regime that holds teachers responsible for student achievement and reflects the increasing accountability placed on teachers to measure student learning (Codd, 2005; Mutch, 2012).

A substantially different conception of assessment was held by teachers in this study that *assessment is for qualifications* which reflects the considerable role they have in assessing students for NCEA qualifications. This finding reinforces an ecological approach whereby teachers' conceptions of assessment are influenced by the assessment environment under which they work (G.T.L. Brown, 2011a; Brown & Michaelides, 2011; Remesal, 2011).

Findings also suggest that New Zealand high school teachers' conceptions of assessment align more with those in high-stakes assessment regimes than those previously reported for their primary colleagues in New Zealand and Australia. This supports Brown and Michaelides' (2011) contention that changes to assessment environments or contexts may lead to different conceptions of assessment and different pathways among those conceptions.

8.2.2 Internal assessment practices

Participants' internal assessment practices were varied, but could generally be described as 'safe' rather than innovative (Hipkins, 2015). Most participants used the Ministry of Education exemplars as a base for designing their own tasks; these were modified to differing extents. Some participants reported extensive modifications to meet the needs and interests of their students, whereas others reported minor modifications to prevent plagiarism. Conditions of assessment varied from in-class tests, with few feedback opportunities, to a more portfolio-orientated approach where formative opportunities were provided. While this tendency of teachers to reproduce the exemplars as their method of assessing the internally-assessed standards has previously been reported by other researchers investigating NCEA (East, 2016; Hume, 2006; Moed, 2010), this finding reinforces this as a general approach to NCEA internal assessment.

Internal assessment practices could be conceived on a continuum from a testing paradigm to an assessment paradigm. However, individual teachers did not necessarily fit neatly into a place on that continuum. Rather, they described a variety of practices that were determined by pragmatics. For example, several justified the use of a portfolio approach in combination with in-class tests because it balanced their workload with students' opportunities to achieve. Previous research has tended to align teachers with *an* assessment identity (Fulmer et al., 2015; Looney et al., 2017; Xu & Brown, 2016) whereas this research has shown teachers can have more than one assessment identity.

Innovative aspects of practice that were revealed included the use of internally-assessed standards to create interesting and motivational courses for the students. This involves using standards from different domains. One popular example was the use of the Business Studies achievement standard 90848, *Carry out, review and refine a business activity within a community context with guidance*, as an authentic opportunity around which to base accounting and economics. Further practices were to discard standards that teachers thought uninteresting, and to maximise the numbers of credits that could be achieved through internal assessment, as opposed to external assessment.

8.2.3 The impact of accountability on internal assessment practices

Participants described a range of influences on their internal assessment practice, but it was the meso-level of school practice, policy and accountability that seemed to dominate. The pressure of school policy and accountability expectations over-rode expressed, preferred internal assessment strategies and also seemed to over-ride teachers' beliefs about assessment. The strongly held conception that the formative function of assessment is foremost was not reconciled with carrying out internal assessment for qualifications. This latter role seems dominated by external and internal policies and accountability. The meso-level of school policy and accountability measures with regards internal assessment seem to be directing teachers actions with regards their internal assessment practices.

8.3 Contributions to the field

8.3.1 Conceptions of assessment are ecologically rational

The use of mixed methods enabled the quantitative data to be considered within the context of the thematic analysis of interviews and revealed converging and complementary associations with four conceptions of assessment. The online survey gathered responses from a greater numbers of teachers than the qualitative interviews alone, and the data from the qualitative interviews were used to explain, reinforce, and elaborate findings from the survey. Both strands of the mixed methods reinforced the view that teachers' conceptions of assessment are ecological (G.T.L. Brown, 2011a; Brown & Michaelides, 2011; Remesal, 2011), because a distinct conception of assessment that relates to participants' role in assessing for NCEA emerged from both the quantitative and qualitative data. In addition, despite this focus on assessing for qualifications, the mixed methods approach revealed a strong affiliation with a formative conception of assessment. The conceptions of assessment of New Zealand secondary school teachers who focus on the senior level of the school have not previously been investigated, and this thesis contends that the strong focus on assessing for qualifications has resulted in a unique conception of assessment for this group of teachers.

8.3.2 Contradictory conceptions are rational

Data from the interviews were also able to explain some seemingly contradictory quantitative findings, such as the simultaneous finding that teachers think the formative function of assessment can be incorporated within internally-assessed achievement standards, but then again it might not. The varying practices among teachers, and of teachers themselves, explains that at times teachers are providing formative opportunities when conducting internal assessment, but at other times and with different standards they are not. To a certain extent this relates to the tensions that teachers face when implementing the formative and summative functions of assessment simultaneously (Irving et al., 2011), but reinforces Newton's (2007) view that teachers attribute intrinsic properties of the assessment process to the terms *formative* and *summative*, rather than seeing *formative* and *summative* as functions of assessment. It also highlights that teachers do not necessarily have one assessment identity, but rather act in different ways at different times (Fulmer et al., 2015; Looney et al., 2017; Xu & Brown, 2016).

Furthermore, an explanation was provided for the strong correlation between the *student accountability* and *improvement* conceptions in the CFA. This high correlation could be due to numerous causes, such as teachers expecting students to be more accountable for their own learning, teachers considering improved assessment results to be an *indication* of improved learning, or aligned with Dann's (2014) explanation of assessment *as* learning under which Dann promotes the prospect of assessment enabling students to understand and gain agency in their own learning. Interview data revealed that the most likely explanation is participants considering improved grades to be an indication of improved learning which supports Torrance's (2007) claim that compliance with assessment criteria can be a proxy for improvement in learning. Participants gave little indication of expecting students to be accountable, neither did they give students much agency in their own learning.

Interview data were also able to explain why teachers can conceive of assessment as having useful formative and summative functions while, at the same time, considering it to have detrimental aspects. Participants shared many examples of the flawed aspects of assessment, such as invalid assessment tasks being provided by the Ministry

of Education and inconsistency in the application of standards, but accepted these as inevitable features of assessment for qualifications. Brown and Michaelides (2011) speculated in their quantitative investigation of Cypriot teachers' conceptions of assessment that teachers consider assessment as indispensable, but at the same time are aware of the misuses. This study supports this speculation.

8.3.3 Vision for NCEA yet to be realised

This research reveals that realisation of the vision for NCEA internal assessment is still some way off. It was envisaged that internal assessment for NCEA would provide students with opportunities to achieve, to work in familiar surroundings, to have ample time to show what they can achieve and to assess a wide range of skills and knowledge (Education Review Office, 2007; NZQA, n.d.e). Hipkins et al. (2016) have also proposed that internal assessment should be enmeshed with learning and that NCEA should provide life-worthy learning experiences for all students. It is contended that moderation procedures and processes, and the use of moderation reports as accountability measures by school managers have encouraged teachers into a cycle of safe, rather than innovative, practice. The vision for internal assessment to be both formative and summative in its function creates dilemmas for schools and teachers, and this research indicates that a greater emphasis is being put on the summative function of internal assessment rather than the learning function. This aligns with Bonner's (2016) conclusion that assessment for summative purposes and assessment for formative purposes create situations of tension and conflict, and that teachers vary in the time and space they allocate to each purpose.

8.3.4 The meso-level is a dominant impact

Teachers' internal assessment practices are influenced by a complex, dynamic range of factors. These include their beliefs about how assessment for qualifications should be carried out, but other important factors are the impact of national and school policies, accountability measures imposed on teachers, students' actions in terms of plagiarism, and teachers' own workload issues. However, a dominant factor influencing teachers' internal assessment practices is the impact of the meso-level of school managers using moderation reports as a measure of accountability. As a result, teachers did not follow one set of practices, but rather used different practices at different times, while always being mindful that their practices had to be acceptable to

their school management. While the summative function of internal assessment remains dominant, and moderation reports and students' achievement results are used as accountability measures, it is unlikely that teachers' internal assessment practices will meet the vision intended.

8.3.5 Teachers are responsible in their role as assessors

Teachers in this research took their role as assessors seriously. Many of the practices and procedures they instigated were to ensure the validity and reliability of internal assessments and the authenticity of students' work. They adopted practices that gave students the maximum opportunity to achieve NCEA, but at the same time reported that they taught the epistemological knowledge of their respective subject. They also used NCEA to motivate and involve students in their own learning and assessment, and as such could be seen to be enacting a socio-cultural approach towards assessment.

8.3.6 Teachers are vulnerable in their role as assessors

Internal assessment for NCEA deprivatises teachers' assessment practices and opens them up to scrutiny. Negative experiences from external moderation had eroded some participants' confidence, led to self-doubt, and increased their vulnerability. While these findings concur with previous research which has reported that moderation procedures have left teachers dispirited (Brooks, 2010; East, 2016; Mizutani, 2009; NZQA, 2014; Office of the Auditor General, 2012; Taylor et al., 2007; Wylie & Bonne, 2016), this study contributes new knowledge that the reactions at the meso-level of the school compound the vulnerability. Accountability measures imposed on top of the NCEA qualifications system have led to feelings of vulnerability when assessing for NCEA. Meeting these accountability measures caused participants stress and concern, which aligns with Nicols' and Harris' (2016) claim that the nature of assessment can affect teachers' well-being, commitment, satisfaction and professional identities. As reported by Parr and Timperley (2016) the degree of autonomy afforded teachers forms "an integral part of teacher identity" (p. 101) and that external accountability can leave teachers feeling undermined. This study supports those claims.

8.4 Limitations of this study

8.4.1 Sampling

One hundred and thirty-five survey participants is considered at the lower end of acceptability for factor analysis (Gorsuch, 1983), and a low response rate is considered a drawback of online questionnaires (Aitken, Power, & Dwyer, 2008). The relatively low number of participants could be an explanation for G.T.L. Brown's (2006) model not being supported. However, after nearly six months of encouraging participation through various accounting and economics teachers' networks, it was considered that all who were willing to take part had done so. In addition, as all participants who volunteered and were available were interviewed, purposive sampling did not take place. This research specifically chose to investigate only accounting and economics teachers and as such does not claim to generalise to the wider population of secondary school teachers.

8.4.2 Instruments

Questionnaires

Questionnaires are not without limitations. Participants can respond to questionnaires in various ways and there is no way of knowing if answers are accurate and truthful (Zohrabi, 2013). In self-administered questionnaires, if misunderstandings arise due to the wording of questions these cannot be clarified and may lead to inaccurate responses. There is a potential that some participants may not have answered accurately and that if the questionnaire had been administered by the researcher, different responses could have been obtained.

Interviews

Interviews are not a neutral tool that elicit rational responses and uncover truths; in fact, there is no guarantee that interviewees are truth tellers (Qu & Dumay, 2011). The interviewer creates an interview situation and the resulting information depends on the characteristics of the interviewer (Denzin & Lincoln, 1998). In semi-structured interviews, the interviewer is required to probe and follow up on questions; these may differ from one participant to another. Also, different interviewers may evoke different responses from the interviewees depending on the traits of the interviewer and the way

questions are asked and probed (Qu & Dumay). Therefore, the data gathered through the semi-structured interviews only relate to the data shared between the interviewer and the interviewees at that particular time.

8.4.3 Analysis

Exploratory factor analysis requires a level of subjectivity because the researcher must make a range of choices and decisions to enhance the quality of the resulting solution (Fabrigar et al., 1999), so the accuracy of the results is dependent on these decisions (Henson & Roberts, 2006; Tabachnick & Fidell, 2007). Choices include the number of factors to retain, type of rotation, and whether to remove items. While the literature was used (e.g., Field, 2013; Osborne et al., 2008; Walkey & Welch, 2010) to guide these decisions, they could still be considered subjective and thus may have an impact on the final outcomes. Naming the factors from EFA can be subjective and was influenced by my understandings of the data and previous research. More testing of the factor structure is required to support this structure, but was beyond the scope of this study. However, according to Beavers et al. (2013), a lack of further testing of a factor structure does not negate its current usefulness and the solution can be meaningful and contribute to research.

Thematic analysis is sometimes critiqued for its simplicity and it relies on the researcher to provide adequate examples from the full data set to persuade readers of the argument (Braun & Clarke, 2006). Naming of the themes is a subjective process, and as with the EFA, those names that I have chosen may not be what another researcher would choose. There is always a level of subjectivity in the interpretation of what participants say. Both inductive and deductive coding was used. For example, the research aimed to find out how teachers conduct internal assessment, so were directly asked this question, and answers coded into the *assessment as practice* theme. Other codes were derived inductively from the data, but these are subject to my “epistemological commitments, and data are not coded in an epistemological vacuum” (Braun & Clarke, p. 12). So while the aim of inductive coding is to code the data without putting it into a pre-existing coding frame, and to allow the themes to be driven by the data, the researcher’s beliefs can influence the final structure.

8.5 Implications of this study

8.5.1 Changes to internal assessment practice are needed

While this research has limitations and further research is needed with New Zealand secondary school teachers, it raises some important issues with regard to the internally-assessed component of NCEA.

The conceptualisation of teachers into the metaphors of ushers and gatekeepers illustrates two overarching paradigms in which teachers had a disposition to situate their internal assessment practice. One paradigm illustrates the potential of internal assessment to meet the needs of a diverse range of students. Key aims of NCEA, and, in particular, the internally-assessment component, were to provide more opportunities for young people to achieve qualifications and to allow a wide range of student performance to be assessed (NZQA & OECD, 2004). This study has shown that these aims can be achieved, with some participants asserting their control over the assessment process to ensure students achieved national qualifications. Klenowski and Wyatt Smith (2014) described how teachers might adapt assessment to cater for students' diverse needs and promote student involvement in a learning community. Participants provided evidence that they attempted a student-centred approach by offering differentiated assessment opportunities and student choice about assessment tasks or topics. Allowing choice and contextualising assessment can also mean that assessment is more culturally responsive because the knowledge and experiences of students are incorporated into learning and assessment, which can be expected to progress social justice goals (Berryman & Bishop, 2016; Klenowski, 2009).

On the other hand, some participants' behaviour reinforced earlier studies on NCEA internal assessment that have shown that teachers tend to implement separate assessment events that strongly imitate the exemplars provided on TKI (East, 2016; Hume, 2006; Moeed, 2010). One purpose of internal assessment is for students to have a form of assessment that is appropriate to them and their context. Widespread usage of the exemplars negates this aim, but is encouraged by NZQA moderators and school management. NZQA needs to consider how much it wants to realise the original vision for internal assessment within NCEA and how this could be achieved.

Currently, teachers do not seem to be encouraged, nor have any incentive to use innovative assessment practices, in fact their good intentions can be thwarted by macro-level and meso-level policy. Teachers need to be empowered, rather than disillusioned, in their role as assessors of internally-assessed standards for NCEA.

8.5.2 Professional development is needed

Professional development can be one source of empowerment and, as Stiggins (2010) claimed, “assessment illiteracy abounds” (p. 233) because too little attention is paid to assessment in initial teacher education and in-service professional development in assessment is infrequent. In addition, professional development on assessment tends to focus on technical aspects of marking or grading and test construction (Campbell, 2013). Best Practice Workshops, which NZQA promote as “support to help with good assessment practice for internally-assessed standards” (NZQA, n.d.k), also seem to be restricted to supporting technical aspects of meeting moderation requirements, making judgements about samples of student work, and interpreting achievement standards. Best Practice Workshops are offered annually and in various locations around New Zealand, so the availability of professional development is not the issue, rather, it is the nature of the professional development available that needs to be addressed.

Current approaches to professional development for internal assessment are not promoting powerful use of internal assessment. If all students are to have a meaningful and useful internal assessment experience, this needs to change. According to Xu and Brown (2016), teacher learning occurs in two main ways: through critical reflective practice (Schön, 1983) or through participation in learning communities (Westheimer, 2008). However, it is possible that neither of these will disrupt teachers’ current practice. Participants in this study who made few modifications to exemplars and set examination-like conditions for assessment accepted this practice as the only viable option. Self-reflection consisted of examining assessment results and, if these were acceptable, few changes were made. Teachers need criteria other than assessment results and moderation agreement rates against which to reflect. Xu and Brown claim that participation in learning communities can engage teachers in professional conversations which may lead them to change their assessment practice, but such communities can also reinforce current practice. Participants provided evidence that

this did occur, such as following a colleague's lead and putting parts of internally-assessed standards into mid-year examinations.

What is needed is for teachers to be challenged on their internal assessment practices, which may entail challenging their beliefs. Although, Guskey (1986) argued that changes in practice follow changes in beliefs, if teachers adopt new practices that are successful, beliefs may also change (Ertmer, 2005). This idea is supported by self-efficacy literature (e.g., Bandura, 1997) whereby successful experiences with change build confidence to make greater change. In terms of changing internal assessment practices this would involve teachers being exposed to successful alternative models because "teachers' practices are unlikely to change without some exposure to what teaching actually looks like when it's being done differently" (Elmore, Peterson, & McCarthey, 1996, p. 241). Teachers providing professional development for other teachers has been used and deemed successful for NCEA internal assessment (Yates, 2012) and could be reinstated. Working with successful colleagues can prompt a perceived need for change as well as assure teachers that changes are possible (Zhao & Cziko, 2001). In addition, more expert teachers can both challenge and support their colleagues during a change process.

Although there is debate about whether beliefs change practice or practice changes beliefs (Guskey, 2006), or whether these are bi-directional (Bandura, 1986) it is clear one cannot be changed without considering the other (Ertmer, 2005). Understanding teachers' conceptions of assessment is integral to changing their assessment practice and professional development facilitators need to be prepared to examine and challenge beliefs. The range of beliefs and practices exhibited by participants shows differentiated professional development is needed for different teachers; a one-size-fits-all approach, such as that currently provided, is unlikely to be successful in affecting change. A model of NCEA professional development such as that trialled in the Senior Subject Advisor pilot and evaluated by Taylor et al. (2007) whereby personalised professional development was provided by teachers, who possessed both disciplinary and NCEA expertise, is more likely to successfully bring about change.

8.5.3 Leadership of internal assessment

School leadership has a central role in the development of teachers' assessment capabilities and leaders, as well as teachers, should participate in professional development (Absolum et al., 2009; Robinson, Hohepa, & Lloyd, 2009; Timperley, Wilson, Barrar & Fung, 2007). School leaders need to examine current assessment practices; they need to resolve conflict about assessment issues; and they need to be able to interpret assessment data to ensure their decisions are well informed (Cowie & Penney, 2016; Scott, Scott, & Webber, 2016). School principals, and other senior managers, need knowledge of effective assessment practice so they can show leadership by encouraging teachers to adopt innovative practices. Some actions (such as moderation reports being used as an accountability measure) of school leaders reported in this study suggested this was not the case. Given the strong influence of the school meso-system on participants' internal assessment practices, change needs to occur at that level for teachers to incorporate innovation into their practice. As Hargreaves (2003) opined, school cultures can be feudal and paternalistic environments that cultivate loyalty and compliance, yet professionals need to be able to disagree, question and critique. School managers need to consider the environments they are cultivating in their schools and whether they are fruitful for students' learning and achievement, or are they encouraging performativity.

Leadership of internal assessment from NZQA also needs to change. NZQA needs to be consistent in its messages to teachers and also to examine its own philosophy towards internal assessment. Hipkins (2013) questions whether the original intent of NCEA is still intact, given that changes such as course endorsements requiring some externally-assessed standards "have arguably made the qualification appear more similar to the previous examination-based system" (p. 19). Hipkins calls for the original intent of NCEA to now be promoted, with which I agree. If internal assessment is conducted as one-off tests, using similar assessment materials, then its aims of meeting the needs of a range of learners are being negated.

Much of the internal assessment material provided by NZQA promotes use of a separate, single event assessment. Changes are needed in NZQA documentation on its website to promote alternative approaches to NCEA internal assessment and the NZQA website also needs to provide consistent information.

NZQA also exerts strong pressure over teachers' beliefs about assessment, beliefs that Rokeach (1968) calls Type D beliefs, that is, those beliefs that are derived from an authority in which we believe. According to Rokeach, these beliefs can be changed if the suggestion for change comes from that authority. If this is the case, NZQA can promulgate a different approach to internal assessment, one that truly does assess students in authentic ways, provides opportunities to show what students can achieve, provides ample time, and gives credit for a wider range of knowledge and skills than the previous school qualification system (Education Review Office, 2007; NZQA, n.d.a, n.d.e). This could start with the way moderators are trained, prepared and deployed.

8.5.4 Teacher education

The findings of this study also have implications for initial teacher education. Teacher educators need to consider that student teachers will enter initial teacher education with a range of experiences and conceptions of assessment. A one-size-fits-all approach is likely to have limited success with student teachers, yet this is the general approach. Student teachers may have naïve understandings about assessment based on their limited experience of being assessed, rather than being assessors. G.T.L Brown (2011c) reported that student teachers did indeed have quite different conceptions of assessment from practising teachers, and concluded that existing conceptions of assessment may create dilemmas for teacher educators in helping student teachers to use assessment in a formative way. Barnes, Fives, and Dacey (2017) explicitly called for student teachers' conceptions of assessment to be considered and for teacher educators to work in tandem with student teachers' belief systems to increase beginning teacher assessment literacy and to use assessment in ways that benefit student learning.

Despite research indicating that teacher educators have little influence on assessment practice due to the short time they spent with student teachers (Deneen & Brown, 2016), teacher educators do have a responsibility to try to affect change. They can support and encourage student teachers to implement NCEA internal assessment innovatively, although student teachers will have varying views about how this should be done based on their own experiences. Teacher educators may also need to consider their own conceptions of assessment and beliefs about how internal assessment should be implemented. To break the practice of using the TKI exemplars as a first reference, teacher educators and student teachers could work together using the NZC and achievement standards to tease out what learning students need to demonstrate and collate ideas about how that could potentially be done. Using a model such as Creative Problem Solving (Isaksen & Treffinger, 1985) to create tasks to assess internally-assessed standards could lead to divergent thinking and innovative practices.

8.5.5 Assessment policy should work for teachers and students

Assessment is complex and multi-faceted, and assessment systems must balance a need to measure and monitor student achievement with the need to improve their learning (Clarke, 2012). This requires consideration of all uses and users of assessment at the policy, school and classroom levels (Cowie & Penney, 2016). Thematic and factor analysis in this study has shown that participants hold a strong view that a key purpose of assessment is to improve student learning, and also to measure their learning through external qualifications. The existence of a conception of assessment that clearly relates to the participants' roles as assessors for NCEA is indicative of the impact that assessment policy has on teachers' conceptions of assessment.

Given that the formative function of assessment is a key support in student learning (Black & Wiliam, 1998), assessment policy should support this function, but Brown and Harris (2009) found that policies that promote the use of assessment data as external accountability mechanisms lead teachers to lose their positive orientation towards the improvement function of assessment. Likewise, Brown and Michaelides (2011) suggested that strong external monitoring polices might not be needed in New Zealand as teachers are already committed to improving the learning of their students. Findings from this study show participants did not endorse the use of assessment results to

measure school quality; therefore, the continuation of such policies is likely to be counter-productive with student learning. As G.T.L. Brown (2008) argued, policy-makers should consider the strong commitment New Zealand teachers have to the improvement conception of assessment and take advantage of this by reducing the accountability stakes associated with assessment data, and rather, use policy to strengthen the formative functions of assessment. The latter emphasis is more likely to lead to genuine improvements in student learning, whereas the former can encourage surface learning and teaching to the test. As Hargreaves (2003) argued, education policy should empower, not constrain teachers, it should be “the wind beneath their wings not the albatross around their necks” (p. 53).

8.6 Areas for future research

This study focussed on teachers of accounting and economics, so an obvious area for further research is to investigate the conceptions of assessment and internal assessment practices of a greater number of secondary teachers which would allow comparisons across subjects and between different groups of teachers. This would enable a bigger-picture view of teachers’ internal assessment practices and whether these are meeting the original aims of NCEA, or rather, are they reverting to imitating the previous examination-based system as Hipkins (2013) claims. A larger sample could support (or not) the factor structure from the exploratory factor analysis and support whether secondary school teachers in New Zealand do have different conceptions of assessment from their primary school colleagues. In particular, is the absorption of a conception of assessment which relates to secondary teachers’ role as assessors of national qualifications widespread?

Further research with NZQA moderators and other NZQA personnel would also be interesting and useful. We need to know their conceptions of assessment and views on how internal assessment should be implemented, also, whether NZQA espoused policy is supported by its actual practice. Moderation reports and feedback from moderators have a powerful influence on teachers’ internal assessment practice and change is unlikely to occur without authority from NZQA.

As the meso-level of school policy wielded a significant influence on teachers' internal assessment practices, further research with school principals and middle managers would be useful to ascertain their beliefs about internal assessment. Are they more focussed on the accountability pressures at the expense of potential learning opportunities? And if this is the case why do they hold such views? School managers are held publically accountable through the Better Public Service Target (New Zealand Government, 2012) of at least 85% of students achieving NCEA Level 2, accountability is also asserted through the media release of NCEA league tables which compare the achievement rates of students and also the associated pressure to recruit students to their schools to ensure financial viability. How do school managers react to such accountability measures and why?

Student voice in relation to internal assessment within NCEA is also relatively rare, and that which exists predates major changes that have occurred in NCEA standards and alignment with the NZC. It would be useful to know what value students place on internal assessment, the feedback they find useful, and conditions of assessment that support learning.

8.7 Significance of the thesis

This study has provided insight into the conceptions of assessment of a group of New Zealand secondary teachers who teach predominantly at the senior level of the school and who have a substantial responsibility in assessing for national qualifications. A major contribution is that the findings from the convergence of the quantitative and qualitative strands show that teachers in this research have absorbed a conception of assessment which relates directly to their role as assessors for a national qualification. In addition, results show that participants' conceptions of assessment align more closely with those of teachers in international high-stakes assessment jurisdictions than they do with their primary colleagues in New Zealand and Australia. These two findings reinforce the ecological approach to teachers' conceptions of assessment whereby these conceptions are influenced by the assessment environment under which they work (G.T.L. Brown, 2011a; Brown & Michaelides, 2011; Remesal, 2011).

This research has also shown that teachers' internal assessment practices for NCEA are influenced by a complex, dynamic range of factors. These include: their beliefs about how assessment for qualifications should be carried out; the impact of macro-level, national and meso-level, school policies; students' actions in terms of plagiarism; and teachers' own workload issues. However, the dominant influence seems to be the accountability pressures exerted at the meso-level of the school. As a result, teachers' internal assessment practices could be described as safe rather than innovative, and dominated by the summative function of assessment of internal assessment, rather than its potential formative function. Participants were responsible in their roles as assessors, but various accountability measures led to feelings of vulnerability.

From this work I have concluded that the vision for NCEA internal assessment is not yet being realised. For this realisation to occur, change is required at a number of levels. This includes a change in approach from NZQA in relation to moderation practices, for school management to encourage rather than (possibly inadvertently) stifle innovative approaches to assessment, as well as professional development for teachers focussed on improving the formative function of their assessment practices.

Good internal assessment can be a powerful tool for improving teaching and learning (Yung, 2000), and through contextualising and allowing choice internal assessment can lead to assessment that is more culturally responsive, which progresses social justice goals (Berryman & Bishop, 2016; Klenowski, 2009). It is important that assessment in the twenty first century maximises these opportunities.

References

- Absolum, M., Flockton, L., Hattie, J., Hipkins, R., & Reid, I. (2009). *Directions for assessment in New Zealand (DANZ): Developing students' assessment capabilities*. Wellington: Ministry of Education.
- Aitken C., Power R., Dwyer R. (2008). A very low response rate in an on-line survey of medical practitioners. *Australian and New Zealand Journal of Public Health*, 32(3), 288–289.
- Adie, L. (2013). The development of teacher assessment identity through participation in online moderation The development of teacher assessment identity through participation in online moderation. *Assessment in Education: Principles, Policy & Practice*, 20(1), 91–106, doi: 10.1080/0969594X.2011.650150
- Adie, L., Klenowski, V., & Wyatt-Smith, C. (2012). Towards an understanding of teacher judgement in the context of social moderation. *Educational Review*, 64(2), 223–240. doi.10.1080/00131911.2011.598919
- Alison, J. (2005). *Teachers talk about NCEA: Research report on focus groups with secondary teachers*. Wellington: Post Primary Teachers Association.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Psychological Association, American Educational Research Association, and National Council on Measurements Used in Education. (1954). Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin*, 51(2), 1–38. doi.org/10.1037/h0053479
- Anderman, E. M., Griesinger, T., & Westerfield, G. (1998). Motivation and cheating during early adolescence. *Journal of Educational Psychology*, 90, 84–93. doi.org/10.1037/0022-0663.90.1.84
- Anderman, L., Freeman, T., & Mueller, C. (2007). The “social” side of social context: Interpersonal and affiliative dimensions of students' experiences and academic dishonesty. In E. Anderman & T. Murdock (Eds.), *Psychology of academic cheating* (pp. 203–228). Burlington, MA: Elsevier Academic Press.

- Aziz, A. (2014). *Indonesian junior secondary school teachers' conceptions of assessment: A mixed methods study* (Doctoral diss.). Victoria University of Wellington, NZ.
- Bachman, L. F., & Purpura, J. E. (2008). Language assessments: Gate-keepers or gate-openers? In B. Spolsky & F. M. Hult (Eds.), *The Handbook of Educational Linguistics* (pp. 456–468). Oxford, UK: Blackwell Publishing.
- Baker, R. (2001, February). *Challenge for educational transformation: Achieving the aim of 'thinking and acting locally, nationally and globally' in a devolved education system*. Paper presented at the International Forum on Education Reforms in the Asia-Pacific Region: Globalisation, Localisation and Individualisation for the Future, The Hong Kong Institute of Education, HKSAR, Hong Kong. Retrieved from <http://www.nzcer.org.nz/research/publications/challenge-educational-transformation-achieving-aim-thinking-and-acting-locally>
- Ball, S. J., Maguire, M., & Braun, A. (2012). *How schools do policy: Policy enactment in the secondary school*. London, England: Routledge.
- Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 93–114). New York, NY: Routledge.
- Bandura, A. (1986). *Social foundations of thought and action*. New Jersey: Prentice Hall.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.
- Barker, A. (1995). Standards-based assessment: The vision and broader factors. In R. Peddie, & B. Tuck (Eds.), *Setting the standards: The assessment of competence in national qualifications* (pp. 15–31). Palmerston North, NZ: Dunmore Press.
- Barnes, N., Fives, H., & Dacey, C. (2017). U.S. teachers' conceptions of the purposes of assessment. *Teaching and Teacher Education*, 65, 107–116.
- Beavers, A., Lounsbury, J., Richards, J., Hucks, S., Skolits, G., & Esquivel, S. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research & Evaluation*, 18(6). Available online: <http://pareonline.net/pdf/v18n6.pdf>
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–246. doi.org/10.1037/0033-2909.107.2.238

- Bentler, P. M., & Chou, C-P. (1987). Practical issues in structural modeling. *Sociological Methods Research*, 16(1), 78–117.
doi.10.1177/0049124187016001004
- Berryman, M. & Bishop, R. (2016). A culturally responsive pedagogy of relations. In D. Fraser & M. Hill (Eds.), *The professional practice of teaching in New Zealand* (pp. 180–197). Australia: Cengage Learning.
- Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2011). Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy and Practice*, 18(4), 451–469.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2002). *Working inside the black box*. London: GL Assessment.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam D. (2003). *Assessment for learning: Putting it into practice*. Maidenhead: Open University Press.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7–74.
doi.10.1080/0969595980050102
- Bogdan, R. C., & Biklen, S. K. (1998). *Qualitative research in education: An introduction to theory and methods* (3rd ed.). Needham Heights, MA: Allyn & Bacon.
- Bonner, S. (2016). Teachers' perceptions about assessment: Competing narratives. In G. T. L. Brown & L. Harris (Eds.), *Handbook of social conditions in assessment* (pp. 21–40). New York: Routledge.
- Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, 36, 81–109. doi.org/10.1017/S0261444803001903
- Borko, H., & Putnam, R. (1996). Learning to teach. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 673–708). New York: Macmillan.
- Bourdieu, P. (1991). *Language and Symbolic Power*. Cambridge: Harvard University Press.
- Bowen, N. K., & Guo, S. (2011). *Structural equation modeling*. New York, NY: Oxford University Press.

- Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. New York, NY: John Wiley.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Broadfoot, P. (2007). *An introduction to assessment*. New York, NY: Continuum International.
- Brookhart, S. (2003). Developing Measurement Theory for Classroom Assessment Purposes and Uses. *Educational Measurement: Issues and Practice*, 22(4), 5–12.
- Brooks, Z. (2010). *Assessing creativity: Teachers' perceptions of the effect of the National Certificate of Educational Achievement (NCEA) on teaching and learning in the secondary drama classroom* (Doctoral diss.). Massey University, Palmerston North, NZ.
- Brown, G. T. L. (2002). *Teachers' conceptions of assessment* (Doctoral diss.). University of Auckland, NZ.
- Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301–318. doi:10.1080/0969594042000304609
- Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged version. *Psychological Reports*, 99, 166–170. doi.org/10.2466/PRO.99.5.166-170
- Brown, G. T. L. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York: Nova Science.
- Brown, G. T. L. (2011a). Teachers' conceptions of assessment: Comparing primary and secondary teachers in New Zealand. *Assessment Matters*, 3, 45–70.
- Brown, G. T. L. (2011b). *An introduction to confirmatory factor analysis and structural equation modelling: Presentation to Research and Development Office, Continuing Professional Development Programme, HKIEd*. Retrieved from <https://www.scribd.com/document/281952342/2011-An-Introduction-to-Confirmatory-Factor-Analysis-CFA-pdf>
- Brown, G. T. L. (2011c). New Zealand prospective teacher conceptions of assessment and academic performance: Neither student nor practicing teacher. In R. Kahn, J. C. McDermott, & A. Akimjak (Eds.), *Democratic access to*

- education* (pp.119–132). Los Angeles, CA: Department of Education. Antioch University Los Angeles.
- Brown, G. T. L., Chaudhry, H., & Dhamija, R. (2015). The impact of an assessment policy upon teachers' self-reported assessment beliefs and practices: A quasi experimental study of Indian teachers in private schools. *International Journal of Educational Research*, *71*, 50–64. doi.org/10.1016/j.ijer.2015.03.001
- Brown, G. T. L., & Gao, L. (2015). Chinese teachers' conceptions of assessment for and of learning: Six competing and complementary purposes. *Cogent Education*, *2*, 1–19. <http://dx.doi.org/10.1080/2331186X.2014.993836>
- Brown, G. T. L., & Harris, L. R. (2009). Unintended consequences of using tests to improve learning: How improvement-oriented resources engender heightened conceptions of assessment as school accountability. *Journal of Multi-Disciplinary Evaluation*, *6*(12), 68–91.
- Brown, G. T. L., Hui, S. K. F., Yu, W. M., & Kennedy, K. J. (2011). Teachers' conceptions of assessment in Chinese contexts: A tripartite model of accountability, improvement, and irrelevance. *International Journal of Educational Research*, *50*, 307–320. doi.10.1016/j.ijer.2011.10.003
- Brown, G. T. L., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for student improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy & Practice*, *16*(3), 347–363. doi.10.1080/09695940903319737
- Brown, G. T. L., & Lake, R. (2006). *Queensland teachers' conceptions of teaching, learning, curriculum and assessment: Comparisons with New Zealand teachers*. Paper presented at Engaging Pedagogies, the Annual Conference of the Australian Association for Research in Education (AARE), Adelaide. Retrieved from <http://www.aare.edu.au/publications-database.php/4974/queensland-teachers-conceptions-of-teaching-learning-curriculum-assessment-comparisons-with-new-zealand>
- Brown, G. T. L., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and Teacher Education*, *27*(1), 210–220. doi.10.1016/j.tate.2010.08.003
- Brown, G. T. L., & Michaelides, M. P. (2011). Ecological rationality in teachers' conceptions of assessment across samples from Cyprus and New Zealand.

- European Journal of Psychology of Education*, 26(93), 319–337.
doi.10.1007/s10212-010-0052-3
- Brown, G. T. L., & Remesal, A. (2012). Prospective teachers' conceptions of assessment: A cross-cultural comparison. *The Spanish Journal of Psychology*, 15(1), 75–89. doi.org/10.5209/rev_SJOP.2012.v15.n1.37286
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press. Retrieved from <http://www.kharazmi-statistics.ir/Uploads/Public/book/Methodology%20in%20the%20Social%20Sciences.pdf>
- Calveric, S. B. (2010). *Elementary teachers' assessment beliefs and practices* (Doctoral diss.). Virginia Commonwealth University, Richmond, VA.
- Campbell, C. (2013). Research on teacher competency in classroom assessment. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp.71–84). Thousand Oaks, CA: Sage.
- Carless, D. (2011). *From testing to productive student learning: Implementing formative assessment in Confucian-Heritage settings*. London: Routledge.
- Chen, J., & Brown, G. T. L. (2013). High-stakes examination preparation that controls teaching: Chinese prospective teachers' conceptions of excellent teaching and assessment. *Journal of Education for Teaching: International Research and Pedagogy*, 39(5), 541–556. doi.10.1080/02607476.2013.836338
- Clarke, M. (2012). *What matters most for student assessment systems: A framework paper*. Washington, DC: The World Bank, The International Bank for Reconstruction and Development.
- Codd, J. (2005). Teachers as 'managed professionals' in the global education industry: The New Zealand experience. *Educational Review*, 57(2), 193–206. doi.10.1080/0013191042000308369
- Codd, J., McAlpine, D., & Poskitt, J. (1995). Assessment policies in New Zealand: Educational reform or political agenda? In R. Peddie & B. Tuck (Eds.), *Setting the standards: The assessment of competence in national qualifications* (pp. 32–59). Palmerston North: Dunmore Press.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). New York: Routledge.
- Comfrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Lawrence Erlbaum.

- Costello, A., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation, 10*(7), 1–9.
- Cowie, B. & Penney, D. (2016) Challenges, tensions and possibilities: An analysis of assessment policy and practice in New Zealand. In S. Scott et al. (Eds.), *Leadership of assessment, inclusion, and learning, the enabling power of assessment 3* (pp. 287–304). Switzerland: Springer Publishing International. doi. 10.1007/978-3-319-23347-5_12.
- Creswell, J. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage.
- Creswell, J. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston: Pearson.
- Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.
- Crombie, W. (1995). Assessing language competencies. In R. Peddie, & B. Tuck (Eds.), *Setting the standards: The assessment of competence in national qualifications* (pp. 99–119). Palmerston North: Dunmore Press.
- da Costa, L., & Remedios, R. (2014). Different methods, different results: Examining the implications of methodological divergence and implicit processes for achievement goal research. *Journal of Mixed Methods Research, 8*(2), 162–179. doi.org/10.1177/1558689813495977
- Dall’Alba, G., & Sandberg, J. (2006). Unveiling professional development: A critical review of stage models. *Review of Educational Research, 76*(3), 383–412. doi.org/10.3102/00346543076003383
- Dann, R. (2014). Assessment as learning: Blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice, 21*(2), 149–166. doi.10.1080/0969594X.2014.898128
- Davison, C. (2004). The contrary culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing, 21*, 305–334.
- Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *Tesol Quarterly, 43*(3), 393–415. doi.org/10.1002/j.1545-7249.2009.tb00242.x

- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self determination in human behavior*. New York: Plenum Press.
- Deneen, C. C., and Brown, G. T. L. (2016). The impact of conceptions of assessment on assessment literacy in a teacher education program. *Cogent Education*, 3, 1225380. doi:10.1080/2331186X.2016.1225380
- Denzin, N. K., & Lincoln, Y. S. (1998). *The landscape of qualitative research: Theories and issues*. Thousand Oaks: Sage.
- Department of Education. (1976). *Towards partnership: The report of the Committee on Secondary Education*. Wellington: Department of Education.
- Department of Education. (1989). *Assessment for better learning: A public discussion document*. Wellington: Author.
- Department of Education. (1986). *Learning and achieving: Second report of the commission of inquiry into curriculum, assessment and qualifications in forms 5 to 7 (CICAQ)*. Wellington, NZ: Author.
- Dewey, J. (1913). *Interest and effort in education*. Cambridge, MA: Riverside Press.
- Dougan, P. (2016, April). NCEA results: How did your school rate? *New Zealand Herald*. Retrieved from http://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=11619826
- East, M. (2016). *Assessing foreign language students' spoken proficiency: Stakeholder perspectives on assessment innovation*. Singapore: Springer.
- Education Counts. (n.d.). *Subject enrolment*. Retrieved from <https://www.educationcounts.govt.nz/statistics/schooling/student-numbers/subject-enrolment>
- Education Review Office. (2007). *The collection and use of assessment information: Good practice in secondary schools*. Retrieved from <http://www.ero.govt.nz/National-Reports/The-Collection-and-Use-of-Assessment-Information-Good-Practice-in-Secondary-Schools-June-2007/Examples-of-good-practice>
- Edwards, F. (2013). Assessing New Zealand high school science: Considerations for teachers' assessment literacy. *Asia-Pacific Forum on Science Learning and Teaching*, 14(2), 1–17.
- Elley, W. (1995). What is wrong with standards-based assessment? In R. Peddie & B. Tuck (Eds.), *Setting the standards: The assessment of competence in national qualifications* (pp. 78–98). Palmerston North: Dunmore Press.

- Elmore, R., Peterson, P., & McCarthy, S. (1996). *Restructuring in the classroom*. San Francisco: Jossey-Bass.
- Ertmer, P. (2005). Teacher pedagogical beliefs: The final frontier in our quest for technology integration? *Educational Technology Research and Development*, 53(4), 25–40.
- Fabrigar, L., Wegener, D., MacCallum, R., & Strahan, E. (1999). Evaluating the use of factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299. doi.org/10.1037/1082-989X.4.3.272
- Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 56–83. doi.10.1080/10705519909540119
- Feilzer, M. Y. (2009). Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research*, 4(1), 6–16. doi.org/10.1177/1558689809349691
- Fetters, M., Curry, L., & Creswell, J. (2013). Achieving integration in mixed methods designs: Principles and practices. *Health Services Research*, 48(6), 2134–2156. doi.10.1111/1475-6773.12117
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage.
- Fitzpatrick, K., & Locke, T. (2008). Sir, do we get credits for this? NCEA, class and ethnicity in New Zealand. In A. St. George, S. Brown, & J. O'Neill (Eds.), *Facing the big questions in teaching: Purpose, power and learning* (pp. 87–98). Melbourne: Cengage Learning.
- Fives, H., & Buehl, M. (2016). Teachers' beliefs, in the context of policy reform. *Policy Insights from the Behavioral and Brain Sciences*, 3(1), 114–121. doi.10.1177/2372732215623554
- Fontana, A., & Frey, J. (1994). Interviewing the art of science. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 361–376). Thousand Oaks, CA: Sage.
- Foucault, M. (1977). *Discipline and Punish: The Birth of the Prison*. New York: Vintage.
- Fullan, M. (2001). *The new meaning of educational change* (3rd ed.). New York: Teachers College, Columbia University.

- Fulmer, G., Lee, I., & Tan, K. (2015). Multi-level model of contextual factors and teachers' assessment practices: An integrative review of research. *Assessment in Education: Principles, Policy & Practice*, 22(4), 475–494.
doi.10.1080/0969594X.2015.1017445
- Gao, M. (2009). Students' voices in school-based assessment of Hong Kong. In D. M. McInerney, G. T. L. Brown, & G. A. D. Liem (Eds.), *Student perspectives on assessment: What students can tell us about assessment for learning* (pp. 107–130). Greenwich: Information Age.
- Gatignon, H. (2014). *Statistical analysis of management data* (3rd ed.). New York: Springer.
- Gebril, A., & Brown, G. T. L. (2013). The effect of high-stakes examination systems on teacher beliefs: Egyptian teachers' conceptions of assessment. *Assessment in Education: Principles, Policy & Practice*, 21(1), 16–33.
doi.10.1080/0969594X.2013.831030
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference*. 11.0 update (4th ed.). Boston: Allyn & Bacon.
- Giluk, T., & Postlethwaite, B. (2015). Big five personality and academic dishonesty: A meta-analytic review. *Personality and Individual Differences*, 72, 59–67.
<http://dx.doi.org/10.1016/j.paid.2014.08.027>
- Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Falmer Press.
- Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory strategies for qualitative research*. London: AldineTransaction. Retrieved from http://www.sxf.uevora.pt/wp-content/uploads/2013/03/Glaser_1967.pdf
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, N.J: Erlbaum.
- Guskey, T. R. (1986). Staff development and the process of teacher change. *Educational Researcher*, 15(5), 5–12.
- Hall, C. (1998). The National Qualifications Framework green paper: What future for the framework? *New Zealand Annual Review of Education*, 7, 29–57.
- Hall, C. (2005). National certificate of educational achievement (NCEA): Is there a third way? In J. Codd & K. Sullivan (Eds.), *Education policy directions in New Zealand* (pp. 154–176). Palmerston North, NZ: Thomson Dunmore Press.
- Hargreaves, A. (1989). *Curriculum and assessment reform*. Toronto: OISE Press.

- Hargreaves, A. (2003). *Teaching in the knowledge society: Education in the age of insecurity*. New York: Teachers College Press.
- Harlen, W. (1994). Towards quality in assessment. In W. Harlen (Ed.), *Enhancing quality in assessment* (pp. 139–145). London: Paul Chapman.
- Harlen, W. (2004). A systematic review of the evidence of the impact on students, teachers and the curriculum of the process of using assessment by teachers for summative purposes. In *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.
- Harlen, W. (2006). On the relationship between assessment for formative and summative purposes. In J. Gardner (Ed.), *Assessment and learning* (pp. 87–102). London: Sage.
- Harlen, W. (2007). *The role of teachers in assessment for learning*. London: Nuffield Foundation. Retrieved from http://www.nuffieldfoundation.org/sites/default/files/assessment_booklet.pdf
- Harris, L., & Brown, G.T. L. (2008, December). *New Zealand teachers' conceptions of the purpose of assessment: Phenomenographic analyses of teachers' thinking*. Paper presented to the Australian Association for Research in Education (AARE) Annual Conference, Brisbane, Australia. Retrieved from <http://www.aare.edu.au/publications-database.php/5659/New-Zealand-teachers'-conceptions-of-the-purpose-of-assessment:-Phenomenographic-analyses-of-teachers'-thinking>
- Harris, L. R., & Brown, G. T. L. (2009). The complexity of teachers' conceptions of assessment: Tensions between the needs of schools and students. *Assessment in Education: Principles, Policy and Practice*, 16(3), 365–381. doi.org/10.1080/09695940903319745
- Hay, P. J. & Macdonald, D. (2008). (Mis) appropriations of criteria and standards-referenced assessment in a performance-based subject. *Assessment in Education: Principles, Policy & Practice*, 15(2), 153-168.
- Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research. *Educational and Psychological Measurement*, 66(3), 393–416.
- Heyvaert, M., Hannes, K., Maes, B., & Onghena, P. (2013). Critical appraisal of mixed methods studies. *Journal of Mixed Methods Research*, 7(4), 302–327. doi.org/10.1177/1558689813479449

- Hipkins, R. (2010). *The evolving NCEA: Findings from the NZCER national survey of secondary schools 2009*. Wellington: New Zealand Council for Educational Research.
- Hipkins, R. (2013). *NCEA one decade on: Views and experiences from the 2012 NZCER National Survey of Secondary Schools*. Wellington: New Zealand Council for Educational Research.
- Hipkins, R. (2015). *Learning to learn in secondary classrooms*. Wellington: New Zealand Council for Educational Research.
- Hipkins, R., & Spiller, L. (2012). *NCEA and curriculum innovation: Learning from change in three schools*. Wellington: New Zealand Council for Educational Research.
- Hipkins, R., Johnston, M., & Sheehan, M. (2016). *NCEA in context*. Wellington: New Zealand Council for Educational Research.
- Hoekstra, A., Brekelmans, M., Beijaard, D., & Korthagen, K. (2009). Experienced teachers' informal learning: Learning activities and changes in behavior and cognition. *Teaching and Teacher Education, 25*, 663–673.
- Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6*(1), 53–60.
- Horsley, J. (2012). Teacher catalysts: Characteristics of teachers who facilitate high academic success. *Australian Journal of Gifted Education, 21*(1), 23–31.
- Hu, L.T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. doi.org/10.1080/10705519909540118
- Hulleman, C., Durik, A., Schweigert, S., & Harackiewicz, J. (2008). Task values, achievement goals, and interest: An integrative analysis. *Journal of Educational Psychology, 2*, 398–416. doi.10.1037/0022-0663.100.2.398
- Hume, A. (2006). *Carrying out a practical investigation: The student-experienced curriculum* (Doctoral diss.). University of Waikato, NZ.
- Hume, A., & Coll, R. K. (2009). Assessment of learning, for learning, and as learning: New Zealand case studies. *Assessment in Education: Principles, Policy & Practice, 16*(3), 269–290, doi.10.1080/09695940903319661
- Hutcheson, G. (1999). Factor analysis. In G. Hutcheson & N. Sofroniu (Eds.), *The multivariate social scientist* (pp. 218–253). London: Sage.

- Irving, S. E., Harris, L. R., & Peterson, E. (2011). 'One assessment doesn't serve all the purposes' or does it? New Zealand teachers describe assessment and feedback. *Asia Pacific Educational Review*, *12*, 413–426.
- Isaksen, D. & Treffinger, D. (1985). *Creative problem solving: The basic course*. Buffalo, NY: Bearly Limited.
- Ivankova, N. (2014). Implementing quality criteria in designing and conducting a sequential QUAN! QUAL mixed methods study of student engagement with learning applied research methods online. *Journal of Mixed Methods Research*, *8*(1), 25–51. doi.org/10.1177/1558689813487945
- Jaba, S. (2013). Acceptance towards school-based assessment among agricultural integrated living skills teachers: Challenges in implementing a holistic assessment. *Journal of Technical Education and Training*, *5*(1), 44–51.
- James, M., & Pedder, D. (2006). Beyond method: Assessment and learning practices and values. *Curriculum Journal*, *17*, 109–138.
doi.10.1080/09585170600792712
- Johnson, R., & Christensen, L. (2014). *Educational research: Quantitative, qualitative and mixed approaches* (5th ed.). Thousand Oaks, CA: Sage.
- Johnson, S. (2011). *A focus on teacher assessment reliability in GCSE and GCE*. Coventry: Office of Qualifications and Examinations Regulation.
- Johnson, S. (2013). On the reliability of high-stakes teacher assessment. *Research Papers in Education*, *16*(1), 91–105. doi.10.1080/02671522.2012.754229
- Joreskog, K. G. (1999). *How large can a standardized coefficient be?* (Unpublished report). Chicago: Scientific Software International. Retrieved from <http://www.ssicentral.com/lisrel/techdocs/HowLargeCanaStandardizedCoefficientbe.pdf>
- Kane, R., & Fontaine, S. (2008). How prepared are New Zealand secondary teachers? Results from a national graduate survey. *New Zealand Journal of Educational Studies*, *43*(2), 29–45.
- Katzer, J., Cook, K. H., & Crouch, W. (1998). *Evaluating information: A guide for users of social science research* (4th ed.). Boston: McGraw Hill.
- Klenowski, V. (2009). Australian indigenous students: Addressing equity issues in assessment. *Teaching Education*, *20*(1), 77–93.
- Klenowski, V., & Wyatt-Smith, C. (2014). *Assessment for education: Standards, judgement and moderation*. London: Sage.

- Kline, R. B. (2010). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.
- Kohn, A. (1993). *Punished by rewards*. Boston: Houghton Mifflin.
- Ledesma, R., & Velero-Mora, P. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. *Practical Assessment, Research & Evaluation*, 12(2). Available online: <http://pareonline.net/getvn.asp?v=12&n=2>
- Lerner, J. S. & Tetlock, P.E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125(2), 255–275.
- Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.
- Linn, R. (1993). Linking results of distinct assessments. *Applied Measurement in Education*, 6(1), 83–102.
- Locker, J., & Cropley, M. (2004). Anxiety, depression and self-esteem in secondary school children: An investigation into the impact of standard assessment tests (SATs) and other important school examinations. *School Psychology International*, 25(3), 333–345. doi.10.1177/0143034304046905
- Looney, A., Cumming, J., van Der Kleij, K., & Harris, K. (2017). Reconceptualising the role of teachers as assessors: Teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*. doi.10.1080/0969594X.2016.1268090
- Mackenzie, N., & Knipe, S. (2006). Research dilemmas: Paradigms, methods and methodology. *Issues in Education Research*, 16(2), 193–205.
- Madaus, G., & Russell, M. (2011). Paradoxes of high-stakes testing. *Journal of Education*, 190, 21–30.
- Madaus, G., Scriven, M., & Stufflebeam, D. (1983). *Evaluation models: Viewpoints on educational and human services evaluation*. Boston: Kluwer-Nijhoff.
- Madjar, I., McKinley, E., Jensen, S., & Van Der Merwe, A. (2009). *Towards university: Navigating NCEA course choices in low-mid decile schools*. Auckland: Starpath Project, The University of Auckland.
- Majid, F. (2011). School-based assessment in Malaysian schools: The concerns of the English teachers. *US-China Education Review*, 8(10), 393–402.
- Matters, G., & Masters, G. (2014). *Redesigning the secondary-tertiary interface: Queensland review of senior assessment and tertiary entrance*. Melbourne: ACER.

- Maxwell, G. (2004, March). *Progressive assessment for learning and certification: Some lessons from school-based assessment in Queensland*. Paper presented at the Third Conference of the Association of Commonwealth Examination and Assessment Boards, Redefining the Roles of Educational Assessment, Nadi, Fiji. Retrieved from https://www.qcaa.qld.edu.au/downloads/publications/paper_progressive_assessment_qld.pdf
- Mertens, D. M. (2005). *Research methods in education and psychology: Integrating diversity with quantitative and qualitative approach* (2nd ed.). Thousand Oaks: Sage.
- Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist*, 30(10), 955–966. doi.org/10.1037/0003-066X.30.10.955
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13–103). Washington, DC: American Council on Education/Macmillan.
- Meyer, L. H., McClure, J., Walkey, F., McKenzie, L., & Weir, K. (2007). *The impact of the NCEA on student motivation: Final report to the Ministry of Education*. Wellington: Ministry of Education and Victoria University.
- Meyer, L. H., McClure, J., Weir, K. F., Walkey, F., & McKenzie, L. (2009). *Motivation and achievement at secondary school: The relationship between NCEA design and student motivation and achievement: A three year follow-up summary report*. Wellington: Ministry of Education and Victoria University.
- Miles, M., & Huberman, M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks: Sage.
- Mills, M., & McGregor, G. (2016). Learning not borrowing from the Queensland education system: Lessons on curricular, pedagogical and assessment reform. *The Curriculum Journal*, 27(1), 113–133, doi.10.1080/09585176.2016.1147969
- Ministry of Education. (n.d.a). *Economics matrix*. Retrieved from <http://ncea.tki.org.nz/Resources-for-Internally-Assessed-Achievement-Standards/Social-sciences/Economics>
- Ministry of Education. (n.d.b). *Accounting matrix*. Retrieved from <http://ncea.tki.org.nz/Resources-for-Internally-Assessed-Achievement-Standards/Social-sciences/Accounting>

- Ministry of Education (1994). *Assessment: Policy to practice*. Wellington: Learning Media.
- Ministry of Education. (1996). *Towards making achievement cool: Achievement in multicultural high schools (AIMHI)*. Retrieved from https://www.educationcounts.govt.nz/publications/maori/english-medium-education/towards_making_achieving_cool_achievement_in_multi_cultural_high_schools_aimhi
- Ministry of Education. (1997). *A future qualifications policy for New Zealand: A plan for the national qualifications framework. Green Paper*. Wellington: Author.
- Ministry of Education. (1999). *The National Qualifications Framework of the future: White paper*. Wellington: Author.
- Ministry of Education. (2005). *Teacher census*. Retrieved from http://www.educationcounts.govt.nz/publications/schooling/teacher_census
- Ministry of Education. (2007). *The New Zealand curriculum*. Wellington: Learning Media.
- Ministry of Education. (2011). *Ministry of Education Position Paper: Assessment (schooling sector)*. Wellington: Crown. Retrieved from assessment.tki.org.nz/content/.../5539/.../MOEAssessmentPositionPaper_October11.pdf
- Ministry of Education. (2017). *Conditions of assessment*. Retrieved from <http://ncea.tki.org.nz/Resources-for-Internally-Assessed-Achievement-Standards/Social-sciences/Accounting/Level-1-Accounting>
- Mizutani, S. (2009). *The mechanism of washback on teaching and learning* (Doctoral diss.). University of Auckland, NZ.
- Mizutani, S., Rubie-Davies, C., Hattie, J., & Philp, J. (2011). Do beliefs about NCEA and its washback effects vary depending on subject? *New Zealand Journal of Educational Studies*, 46(2), 47–59.
- Moeed, H. (A). (2010). *Science investigation in New Zealand secondary schools: Exploring the links between learning, motivation and internal assessment in year 11* (Doctoral diss.). Victoria University of Wellington, NZ.
- Morgan, D. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, 1(1), 48–76. doi.org/10.1177/2345678906292462

- Muijs, D. (2011). *Doing quantitative research with SPSS*. London: Sage.
- Muñoz, A. P., Palacio, M., & Escobar, L. (2012). Teachers' beliefs about assessment in an EFL context in Colombia. *PROFILE*, *14*, 143–158.
- Murdock, T. B., Hale, N. M., & Weber, M. J. (2001). Predictors of cheating among early adolescents: Academic and social motivations. *Contemporary Educational Psychology*, *26*, 96–115. doi.org/10.1006/ceps.2000.1046
- Murdock, T., Stephens, J., & Grotewiel, M. (2016). Student dishonesty in the face of assessment: Who, why, and what we can do about it. In G. T. L. Brown & L. Harris (Eds.), *Handbook of social conditions in assessment* (pp. 187–203). New York: Routledge.
- Mutch, C. (2012). Assessment *for, of and as* learning: Developing a sustainable assessment culture in New Zealand schools. *Policy Futures in Education*, *10*(4), 374–385. doi.org/10.2304/pfie.2012.10.4.374
- Nagy, P. (2000). The Three Roles of Assessment: Gatekeeping, Accountability, and Instructional Diagnosis. *Canadian Journal of Education*, *25*(4), 262–279.
- New Zealand Commerce and Economics Teachers Association. (2006). *Submission to the Ministry of Education on the Draft New Zealand Curriculum*. Retrieved from <https://nzcurriculum.tki.org.nz/content/download/725/4515/file/sec72.pdf>
- New Zealand Government. (2012). *Delivering better public services: Boosting skills and employment by increasing educational achievement for young people*. Wellington: Author. Retrieved from <https://education.govt.nz/assets/Documents/Ministry/BPS/BPSYoungPeopleWEB.pdf>
- New Zealand Post Primary Teachers' Association. (1969). *Education in change: Report of the curriculum review group*. Auckland: Longman Paul.
- New Zealand Post Primary Teachers Association. (2007, September). *Nothing endures but change*. Paper presented to PPTA Annual Conference. Retrieved from <http://ppta.org.nz/publication-library/annual-conference-papers/>
- New Zealand Post Primary Teachers Association. (2015, April). *The NCEA: Can it be saved?* Paper presented to PPTA Annual Conference. Retrieved from <http://ppta.org.nz/publication-library/annual-conference-papers/>
- New Zealand Post Primary Teachers Association. (2016). *PPTA Workload Taskforce Report: Report of the 2015 investigation into issues of workload intensification*

- for secondary school teachers in New Zealand*. Retrieved from <http://ppta.org.nz/advice-and-issues/teacher-workload/>
- New Zealand Qualifications Authority & Organization for Economic Cooperation and Development. (2004). *The role of national qualifications in promoting lifelong learning: Background report for New Zealand*. Paris: OECD.
- New Zealand Qualifications Authority. (1991). *Designing the framework: A discussion document about restructuring national qualifications*. Wellington: NZQA. Retrieved from <http://files.eric.ed.gov/fulltext/ED354327.pdf>
- New Zealand Qualifications Authority. (1992). *Learning to learn: An introduction to the National Qualifications Framework*. Wellington: NZQA. Retrieved from <http://www.voced.edu.au/content/ngv%3A20914>
- New Zealand Qualifications Authority. (2011). *Guidelines for assessing Level 1 literacy and numeracy unit standards*. Retrieved from <http://www.nzqa.govt.nz/assets/qualifications-and-standards/qualifications/ncea/NCEA-subject-resources/Literacy-and-Numeracy/Resources/Planning-implementation-and-assesment/Guidelines-for-assessing-Literacy-and-Numeracy-unit-standardsJan13.pdf>
- New Zealand Qualifications Authority. (2014). *An independent review of the effectiveness of NZQA's implementation of the 2007 NCEA enhancements: Report of the panel*. Retrieved from <http://www.nzqa.govt.nz/assets/qualifications-and-standards/qualifications/ncea/Report-of-the-NCEA-enhancements-review.pdf>
- New Zealand Qualifications Authority. (2016). *Annual report on NCEA and New Zealand scholarship data and statistics (2015)*. Wellington: Author.
- New Zealand Qualifications Authority. (2017a, March). *Initial teacher educators: Quality assurance*. Powerpoint presentation to initial teacher education student teachers, Victoria University, School of Education.
- New Zealand Qualifications Authority. (2017b). *QA News Issue 95–31*. Retrieved from <http://www.nzqa.govt.nz/about-us/publications/qa-news/issue-95/#NCEAresultsrelease>
- New Zealand Qualifications Authority. (n.d.a). *The National Certificate of Educational Achievement: An introduction for parents and students*. Wellington. (not dated but thought to be published in 2000).

- New Zealand Qualifications Authority. (n.d.b). *How NCEA works*. Retrieved from <http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/understanding-ncea/how-ncea-works>
- New Zealand Qualifications Authority. (n.d.c). *Fact Sheet #2: NCEA and education in New Zealand*. Retrieved from <http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/understanding-ncea/the-facts/factsheet-2/>
- New Zealand Qualifications Authority. (n.d.d). *Factsheet # 3: How does NCEA's standards-based system challenge, motivate and reward students?* Retrieved from <http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/understanding-ncea/the-facts/factsheet-3>
- New Zealand Qualifications Authority. (n.d.e). *NCEA endorsements*. Retrieved from <http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/understanding-ncea/how-ncea-works/endorsements/>
- New Zealand Qualifications Authority. (n.d.f). *Factsheet # 8: Why are some achievement standards externally assessed and some internally assessed?* Retrieved from <http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/understanding-ncea/the-facts/factsheet-8/>
- New Zealand Qualifications Authority. (n.d.g). *Where did NCEA come from?* Retrieved from <http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/understanding-ncea/history-of-ncea/where-did-ncea-come-from/>
- New Zealand Qualifications Authority. (n.d. h). *External moderation*. Retrieved from <http://www.nzqa.govt.nz/providers-partners/assessment-and-moderation/managing-national-assessment-in-schools/secondary-moderation/external-moderation/>
- New Zealand Qualifications Authority. (n.d.i). *Assessment opportunities in schools*. Retrieved from <http://www.nzqa.govt.nz/providers-partners/assessment-and-moderation/assessment-of-standards/generic-resources/gathering-evidence-of-achievement/assessment-opportunities-in-schools/>
- New Zealand Qualifications Authority. (n.d j). *Gathering evidence of achievement*. Retrieved from <http://www.nzqa.govt.nz/providers-partners/assessment-and-moderation/assessment-of-standards/generic-resources/gathering-evidence-of-achievement/>

- New Zealand Qualifications Authority. (n.d.k). *Best practice workshops and assessor support: for internal standards*. Retrieved from <http://www.nzqa.govt.nz/about-us/events/assessor-support/>
- Newton, P. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 149–170, doi.10.1080/09695940701478321
- Newton, P., & Shaw, S. (2014). *Validity in educational and psychological assessment*. London: Sage.
- Nicholls, J. G. (1979). Quality and equality in intellectual development: The role of motivation in education. *American Psychologist*, 34, 1071–1084. doi.org/10.1037/0003-066X.34.11.1071
- Nichols, S., & Harris, L. (2016). Accountability assessment effects on teachers and schools. In G. T. L. Brown & L. Harris (Eds.), *Handbook of social conditions in assessment* (pp. 111–129). New York: Routledge.
- Noble, A. J., & Smith, M. L. (1994). *Old and new beliefs about measurement-driven reform: 'The more things change, the more they stay the same'* (Tech. Rep. No. 373). Tempe, AZ: Center for the Study of Evaluation, Arizona State University.
- Nusche, D., Laveault, D., MacBeath, J., & Santiago, P. (2012). *OECD reviews of evaluation and assessment in education: New Zealand 2011*. Paris: OECD. <http://dx.doi.org/10.1787/9789264116917-en>
- O’Cathain, A., Murphy, E., & Nicoll, J., (2010). Three techniques for integrating data in mixed methods studies. *BMJ*, 341, c4587. Retrieved from <http://www.bmj.com/content/341/bmj.c4587.full>
- O’Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer’s MAP test. *Behavior Research Methods, Instrumentation, and Computers*, 32(3), 396–402.
- Office of the Auditor General. (2012). *New Zealand Qualifications Authority: Assuring the consistency and quality of internal assessment for NCEA*. Wellington: Author.
- Osborne, J. W., Costello, A. B., & Kellow, J. T. (2008). Best practices in exploratory factor analysis. In J. W. Osborne (Ed.), *Best practices in quantitative methods* (pp. 205–213). Thousand Oaks, CA: Sage.

- Pajares, M. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307–332.
doi.org/10.3102/00346543062003307
- Parr, J., & Timperley, H. (2016). Section discussion: Teachers and assessment. Enhancing assessment capability. In G. T. L. Brown & L. Harris (Eds.), *Handbook of social conditions in assessment* (pp. 21–40). New York: Routledge.
- Patton, M. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.
- Pedder, D., & Opfer, D. (2013). Professional learning orientations: Patterns of dissonance and alignment between teachers' values and practices. *Research Papers in Education*, 28(5), 539–570. doi.org/10.1080/02671522.2012.706632
- Pedulla, J., Abrams, L., Madaus, G. F., Russell, M., Ramos, M., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Chestnut Hill, MA: National Board on Educational Testing and Public Policy, Boston College. Retrieved from <http://www.bc.edu/research/nbetpp/>
- Pishghadam, R., & Shayesteh, S. (2012). Conceptions of assessment among Iranian EFL teachers. *The Iranian EFL Journal*, 8, 9–23.
- Pryor, J. & Crossouard, B. (2008). A socio-cultural theorisation of formative assessment. *Oxford Review of Education*, 34(1), 1–20. DOI: 10.1080/03054980701476386
- Punch, K. F. (2005). *Introduction to social research: Quantitative and qualitative approaches*. Thousand Oaks, CA: Sage.
- Qian, D. (2014). School-based English language assessment as a high stakes examination component in Hong Kong: Insights of frontline assessors. *Assessment in Education: Principles, Policy & Practice*, 21(3), 251–270. doi.10.1080/0969594X.2014.915207
- Qu, S., & Dumay, J. (2011). The qualitative research interview. *Qualitative Research in Accounting & Management*, 8(3), 238–264. doi.org/10.1108/11766091111162070
- Rabi, S. M., Patton, L. R., Fjortoft, N., & Zgarrick, D. P. (2006). Characteristics, prevalence, attitudes, and perceptions of academic dishonesty among pharmacy

- students. *American Journal of Pharmaceutical Education*, 70(4), 73–81.
doi.org/10.5688/aj700473
- Raman, K., & Yamat, H. (2014). English teachers' voices on the challenges of the school-based assessment. *Frontiers of Language and Teaching*, 5, 66–74.
- Rata, E., & Taylor, A. (2015). Knowledge equivalence discourse in New Zealand secondary school science. *New Zealand Journal of Education Studies*, 50, 223–238. doi.10.1007/s40841-015-0020-1
- Rawlins, P. (2008). Unlocking the formative potential of NCEA. *New Zealand Journal of Teachers' Work*, 5(2), 105–118.
- Remesal, A. (2011). Primary and secondary teachers' conceptions of assessment: A qualitative study. *Teaching and Teacher Education*, 27(2), 472–482.
<http://dx.doi.org/10.1016/j.tate.2010.09.017>
- Richardson, V., & Placier, P. (2001). Teacher change. In V. Richardson, *Handbook of research on teaching* (4th ed., pp. 905–950). Washington, D. C.: American Education Research Association.
- Riley, B. (2014). *Science, data and decisions in New Zealand's education system*. Wellington: Fulbright New Zealand. Retrieved from
https://www.fulbright.org.nz/wp-content/uploads/2014/08/axford2014_riley.pdf
- Robinson, V., Hohepa, M., & Lloyd, C. (2009). *School leadership and student outcomes: Identifying what works and why*. Auckland, NZ: New Zealand Ministry of Education. Retrieved from:
http://www.educationcounts.govt.nz/__data/assets/pdf_file/0015/60180/BESLeadership-Web.pdf
- Rokeach, M. (1968). A theory of organisation and change within value-attitude systems. *Journal of Social Issues*, XXIV(1), 13–33. doi.org/10.1111/j.1540-4560.1968.tb01466.x
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144. doi.org/10.1007/BF00117714
- Sannino, A. (2008). Sustaining a non-dominant activity in school: Only a utopia? *Journal of Educational Change*, 9(4), 329–338. doi.10.1007/s10833-008-9080-z
- Schön, D. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.

- Scott, S., Scott, D. E., & Webber, C. F. (2016). The assessment KSA learning journey: Expanding the 4L – Life-Long Learning Leader – Framework. In S. Scott et al. (Eds.), *Leadership of assessment, inclusion, and learning, the enabling power of assessment 3* (pp. 201–230). Switzerland: Springer. doi.10.1007/978-3-319-23347-5_12.
- Scriven, M. (1966). *The methodology of evaluation*. Social Science education consortium, Publication 100. Retrieved from <http://www.comp.dit.ie/dgordon/Courses/ILT/ILT0005/TheMethodologyOfEvaluation.pdf>
- Segars, M., & Tillema, H. (2011). How do Dutch secondary teachers and students conceive the purpose of assessment? *Studies in Educational Evaluation, 37*, 49–54. doi.org/10.1016/j.stueduc.2011.03.008
- Senge, P. M. (1990). *The fifth discipline: The art & practice of the learning organization*. New York: Doubleday Business.
- Sheehan, M. (2013). “Better to do than receive”: *Learning to think historically through internally assessed course work*. Wellington, NZ: Teaching and Learning Research Initiative.
- Shohamy, E. (2001). *The Power of Tests: A Critical Perspective on the Uses of Language Tests*. Essex: Longman.
- Shulruf, B., Hattie, J., & Tumen, S. (2010) New Zealand’s standard based assessment for secondary schools (NCEA): implications for policy makers. *Asia Pacific Journal of Education (30)2*, 141–165. doi.10.1080/02188791003721598
- Sivo, S., Fan, X., Witta, E. L., & Willse, J. T. (2006). The search for “optimal” cutoff properties: Fit index criteria in structural equation modeling. *The Journal of Experimental Education, 74(3)*, 267–288. doi.10.3200/JEXE.74.3.267-288
- Stanley, G., MacCann, R., Gardner, J., Reynold, L., & Wild, I. (2009). *Review of teacher assessment: What works best and issues for development*. UK: Qualifications and Curriculum Authority.
- Starkey, L., Stevens, S., Taylor, M., Toia, R., Yates, A., Hall, C., McKenzie, L., & Meyer, L. H. (2006). *School/cluster based secondary qualifications professional development: Review of secondary schools’ use of NCEA*

- professional development resources, 2005-2006*. Wellington: Ministry of Education.
- Starkey, L., Yates, A., Meyer, L. H., Hall, C., Taylor, M., Stevens, S., & Toia, R. (2009). Professional development design: Embedding educational reform in New Zealand. *Teaching and Teacher Education*, 25(1), 181–189. doi.org/10.1016/j.tate.2008.08.007
- Steiger, J. H. (1990). Structural model evaluation and modification. *Multivariate Behavioral Research*, 25, 214–212. doi.org/10.1207/s15327906mbr2502_4
- Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modelling. *Personality and Individual Differences*, 42(5), 893–898. doi.org/10.1016/j.paid.2006.09.017
- Stewart, S., Gray, S., & Pilcher, E. (2007). *Assessment practices of Level 3 NCEA teachers (2004-2006)*. Wellington: New Zealand Qualifications Authority.
- Stiggins, R. (1995). Assessment literacy for the 21st century. *The Phi Delta Kappan*, 77(3), 238–245.
- Stiggins, R. J. (2010). Essential formative assessment competencies for teachers and school leaders. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of Formative Assessment* (pp. 233–250). New York: Routledge.
- Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. Abingdon, UK: Routledge.
- Strauf, R. (2016). *Mixed methods: Interviews, surveys and cross cultural comparisons*. Cambridge, UK: Cambridge University Press.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). New York: Allyn & Bacon.
- Taras, M. (2010). Assessment for learning: Assessing the theory and evidence. *Social and Behavioral Sciences*, 2, 3015–3022. doi.org/10.1016/j.sbspro.2010.03.457
- Tashakkori, A., & Teddlie, C. (2003). The past and future of mixed methods research: From data triangulation to mixed model designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioural research* (pp. 671–701). Thousand Oaks, CA: Sage.
- Taylor, M., Kinsella, P., Yates, A., McKenzie, L. H., & Meyer, L. (2007). *Evaluation of the Senior Subject Adviser Pilot Initiative 2007*. Wellington: Ministry of Education.

- Taylor, M., Yates, A., Meyer, L. H., & Kinsella, P. (2011). Teacher professional leadership in support of teacher professional development. *Teaching and Teacher Education*, 27(1), 85–94. doi.org/10.1016/j.tate.2010.07.005
- Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research*. Thousand Oaks, CA: Sage.
- Teddlie, C., & Tashakkori, A. (2010). Overview of contemporary issues in mixed methods research. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (2nd ed.). Thousand Oaks, CA: Sage.
- Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127–146). New York: Macmillan.
- Thrupp, M. (2017). New Zealand's national standards policy: How should we view it a decade on? *Teachers and Curriculum*, 17(1), 11–15.
- Thrupp, M., & Alcorn, N. (2010). A little knowledge being a dangerous thing? Decile-based approaches to developing NCEA league tables. *New Zealand Annual Review of Education*, 20, 52–73.
- Timperley, H. S., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher Professional Learning and Development Best Evidence Synthesis*. Wellington: Ministry of Education.
- Tong, S. Y. A., & Adamson, B. (2015). Student voices in school-based assessment. *Australian Journal of Teacher Education*, 40(2), 15–28. doi.org/10.14221/ajte.2015v40n2.2
- Torrance, H. (1995). Introduction. In H. Torrance (Ed.), *Evaluating authentic assessment: Problems and possibilities in new approaches to assessment* (pp. 1–8). Buckingham: Open University Press.
- Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education: Principles, Policy & Practice*, 14(3), 281–294. doi.10.1080/09695940701591867
- Turner, T. R., Irving, S. E., Li, M., & Yuan, J. (2010). *Availability of NCEA standards: Impact on success rate*. Auckland: University of Auckland, Starpath Project.
- Vygotsky, L. (1978). *Mind and society*. Cambridge, MA: Harvard University Press.

- Walkey, F., & Welch, G. (2010). *Demystifying factor analysis: How it works and how to use it*. USA: Xlibris.
- Walliman, N. (2005). *Your research project* (2nd ed.). London: Sage.
- Weiner, B. (1985). Spontaneous causal thinking. *Psychological Bulletin*, 97(1), 74–84. doi:10.1037/0033-2909.97.1.74
- Westheimer, J. (2008). Learning among colleagues: Teacher community and the shared enterprise of education. In M. Cochran-Smith, S. Feiman-Nemser, & D. J. McIntyre (Eds.), *Handbook of research on teacher education: Enduring questions in changing contexts* (pp. 756–783). New York, NY: Routledge.
- Wiliam, D. (1994). *Towards a philosophy for educational assessment*. Update on a paper given at the British Educational Research Association's 20th annual conference, Oxford. Retrieved from <https://pdfs.semanticscholar.org/8025/955a8c637001fa4d467869a5695214036e78.pdf>
- Wiliam, D. (2008). Quality in assessment. In S. Swaffield (Ed.), *Unlocking assessment: Understanding for reflection and application*. Abingdon: Routledge.
- Willig, C., & Stainton-Rogers, W. (2008). *Handbook of qualitative research in psychology*. London: Sage.
- Wolf, D., Bixby, J., Glenn, J., & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. *Review of Research in Education*, 17, 31-74.
- Wyatt-Smith, C. M., Klenowski, V., & Gunn, S. (2010). The centrality of teachers' judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, Policy & Practice*, 17(1), 59–75. doi.org/10.1080/09695940903565610
- Wylie, C. (1999). *Ten years on: How schools view educational reform*. Wellington, NZ: New Zealand Council of Educational Research.
- Wylie, C. (2012). *Vital connections: Why we need more than self-managing schools*. Wellington, NZ: New Zealand Council of Educational Research.
- Wylie, C., & Bonne, L. (2016). *Secondary schools in 2015: Findings from the NZCER national survey*. Wellington, NZ: New Zealand Council for Educational Research. Retrieved from <http://dx.doi.org/10.18296/rep.0001>

- Wyse, D., & Torrance, H. (2009). The development and consequences of national curriculum assessment for primary education in England. *Educational Research, 51*(2), 213–238. doi.org/10.1080/00131880902891479
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education, 58*, 149–162. <https://doi.org/10.1016/j.tate.2016.05.010>
- Yan, Z. (2014). Predicting teachers' intentions to implement school-based assessment using the theory of planned behaviour. *Educational Research and Evaluation: An International Journal on Theory and Practice, 20*(2), 83–97. doi.10.1080/13803611.2013.877394
- Yates, A. (2012). Devolving professional learning to subject specialists: Towards embedding assessment reform in New Zealand. *Professional Development in Education, 38*(4), 613–629.
- Yip, D.Y., & Cheung, D. (2005). Teachers' concerns on school-based assessment practical work. *Journal of Biological Education, 39*(4), 156–162. doi.org/10.1080/00219266.2005.9655989
- Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology, 9*(2), 79–94. doi.org/10.20982/tqmp.09.2.p079
- Yung, B. H. W. (2000). *Teachers' beliefs and their teaching of practical work in a school-based assessment scheme* (Doctoral diss.). University of Hong Kong, Hong Kong.
- Yung, B. H. W. (2012). Issues and challenges in school-based assessment of science practical work. In K. C. D. Tan & M. Kim (Eds.), *Issues and challenges in science education research: Moving forward* (pp. 125–140). Netherlands: Springer.
- Zhang, Y., & Wildemuth, B. M. (2009). Qualitative analysis of content. In B. Wildemuth (Ed.), *Applications of social research methods to questions in information and library science* (pp. 318–329). Santa Barbara, CA: Libraries Unlimited.
- Zhao, Y., & Cziko, G. A. (2001). Teacher adoption of technology: A perceptual control theory perspective. *Journal of Technology and Teacher Education, 9*(1), 5–30.

- Zohrabi, M. (2013). Mixed method research: Instruments, validity, reliability and reporting findings. *Theory and Practice in Language Studies*, 3(2), 254–262.
- Zwick, W. R. & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99(3), 432–442.

Appendix 1: The survey

Conceptions of assessment

Thank you for completing this questionnaire. The questionnaire is in three parts and should take no longer than 20 minutes to complete.

Part A is designed to find out about your beliefs and understandings about ASSESSMENT, whatever that term means to you. Please answer the questions using your own understanding of assessment. There are no right or wrong answers; I am interested only in your frank opinions. Please give your initial response to the questions, rather than deliberating over them

Part B is designed to find out your knowledge in relation to NCEA internal assessment. Again there are no right or wrong answers.

Part C is designed to collect demographic data relating to you.

Thank you very much for answering these questions.

Anne Yates, PhD student, Victoria University of Wellington.

Please continue ...

Please tick one box for each statement

Part A: Conceptions of Assessment	Strongly Disagree	Mostly Disagree	Slightly Agree	Moderately Agree	Mostly Agree	Strongly Agree
1. Assessment provides information on how well schools are doing	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Assessment places students into categories	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Assessment is a way to determine how much students have learned from teaching	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Assessment provides feedback to students about their performance	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Assessment is integrated with teaching practice	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Assessment results are trustworthy	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Assessment forces teachers to teach in a way against their beliefs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Teachers conduct assessments but make little use of the results	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Assessment results should be treated cautiously because of measurement error	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Assessment is an accurate indicator of a school's quality	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Assessment is assigning a grade or level to student work	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. Assessment establishes what students have learned	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. Assessment feeds back to students their learning needs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. Assessment information modifies ongoing teaching of students	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Please tick one box for each statement

Part A (cont): Conceptions of Assessment	Strongly Disagree	Mostly Disagree	Slightly Agree	Moderately Agree	Mostly Agree	Strongly Agree
15. Assessment results are consistent	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16. Assessment is unfair to students	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17. Assessment results are filed & ignored	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18. Teachers should take into account the error and imprecision in all assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19. Assessment is a good way to evaluate a school	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20. Assessment determines if students meet qualifications standards	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21. Assessment measures students' higher order thinking skills	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
22. Assessment helps students improve their learning	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
23. Assessment allows different students to get different instruction	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
24. Assessment results can be depended on	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25. Assessment interferes with teaching	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
26. Assessment has little impact on teaching	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
27. Assessment is an imprecise process	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

CoA-III Abridged ©2002, Dr. Gavin Brown, University of Auckland

Please continue ...

Please tick one box for each statement

Part B: Knowledge of NCEA internal assessment	Strongly Disagree	Mostly Disagree	Slightly Agree	Moderately Agree	Mostly Agree	Strongly Agree
1 I am knowledgeable in motivating all senior students to do their best (not just high achievers)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2 I am knowledgeable generating student interest in my senior subject	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3 I am knowledgeable making appropriate judgment decisions against NCEA standards, especially around the grade margins	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4 I am knowledgeable adapting web material tasks to meet the requirements of an internally assessed standard	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5 I am knowledgeable writing NCEA internal assessment tasks and assessment schedules	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6 I am knowledgeable making judgment decisions for NCEA assessment, especially around the grade margins	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7 I am knowledgeable in creating a Level 2 accounting and/or economics course that meets the needs of my particular students.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Part C - Would you also provide the following information?

1) What is your role? (Tick one only)

- Teacher
- MU holder
- Head of Department
- Assistant or Deputy Principal
- Principal
- Other:

2) What is your highest degree? (Tick one only)

- Bachelor
- Postgraduate Certificate
- Postgraduate Diploma
- Master
- Doctor

3) What is your specialist teaching subject?

- Economics
- Accounting
- Both

4) For how many years have you taught in New Zealand? (*Tick one only*)

- Less than 2
- Between 2 and 5
- Between 6 and 10
- Between 11 and 20
- More than 20

5) Are you an overseas trained teacher?

- Yes
- No

6) What training in educational assessment have you had? (*Tick all that apply*)

- None
- Some hours as part of pre-service training
- Workshops or Seminars
- Completed an undergraduate paper
- Completed a postgraduate paper
- Other: (*give details*)

7) In accounting, which internally assessed achievement standards do your students complete? (*Tick all that apply*)

- AS91481 (2.5)
- AS91179 (2.6)
- AS91386 (2.7)

For each standard that the students complete please identify the tasks used to assess the standards, i.e. Own, Commercial, TKI exemplars, modified TKI exemplar, other (give details).

Please explain why you use these tasks.

8) In economics, which internally assessed achievement standards do your students complete? (Tick all that apply)

- AS91225 (2.4)
- AS91226 (2.5)
- AS91227 (2.6)
- AS91228 (2.7)

For each standard that the students complete please identify the tasks used to assess the standards, i.e. Own, Commercial, TKI exemplars, modified TKI exemplar, other (give details).

Please explain why you use these tasks.

9) Please provide the range of grades your students achieved in 2012 for each of the internally assessed standards.

Accounting

- AS91481 (2.5): N , A , M , E
- AS91179 (2.6): N , A , M , E
- AS91386 (2.7) N , A , M , E

Economics

- AS91225 (2.4): N , A , M , E
- AS91226 (2.5): N , A , M , E
- AS91227 (2.6): N , A , M , E
- AS91228 (2.7): N , A , M , E

10) What is your gender? (Tick one only)

- Female
- Male

11) Which ethnic group do you belong to? (Tick one only)

- NZ European / Pākehā
- NZ Māori
- Pacific Nation
- Asian: (give details) _____
- Other: (give details) _____

12) At which type of school do you teach? (Tick one only)

- State co-educational secondary school
- State single sex secondary school
- Integrated secondary school
- Private secondary school

13) The decile rating of my school is _____

14) The researcher would really like to hear your thoughts about NCEA internal assessment and would appreciate your participation in an interview. Please indicate whether you are willing to volunteer to be interviewed as part of this study

- Yes
- No

If the answer to Q15 is yes, please provide the following

14) What is your school's name? _____

16) What is your name? _____

17) What is your email address _____

Appendix 2: the generic interview schedule

1. What do you think is the purpose of assessment in schools?
2. Do you think the purpose varies between junior and senior secondary school?
3. What do you think is the main purpose of formative assessment?
4. What do you think is the main purpose of summative assessment?
5. Do you think it is possible to use NCEA in a formative way? Please explain.
6. Do you use the TKI exemplars to assess the internally assessed standards? Why? Why not?
7. Are there any other approaches you would be willing to use for assessing these standards?
8. NZQA information about NCEA internal assessment states that the internal assessment component of NCEA allows schools to assess in authentic ways and provides students with the opportunity to show what they can achieve, to work in familiar surroundings and to have ample time to show what they can achieve.
 - a. Are you able to achieve this?
 - b. If answer is yes – please provide me with an example
 - c. What makes this assessment authentic?
9. NZQA also claims the ability to internally assess and improved flexibility in programme design gives schools opportunities to develop specific programmes and assessment regimes that suit their students. Do you do this?
 - a. If answer is yes - Please provide me with an example.
 - b. If answer is no - Why don't you think you are able to achieve this?

Appendix 3 Ethics approval



FACULTY OF EDUCATION
PO Box 17-310 Wellington 6012, New Zealand
Website www.victoria.ac.nz

26 March 2012

Anne Yates
Lecturer
Victoria University of Wellington Faculty of Education
C/- School of Educational Psychology and Pedagogy
Donald Street
Wellington

Dear Anne

RE: Ethics application SEPI/2013/10 RM RM 19669

I am pleased to advise you that your ethics application '**An investigation of the relationships between teachers' beliefs and practices about assessment and students' achievement in NCEA.**', with the required changes, has been approved by the Victoria University of Wellington Faculty of Education Ethics Committee. Please note that the approval for your research to commence is from the date of this letter.

Best wishes for your research.

Yours Sincerely



Dr Sue Cornforth

Co-Convener
Victoria University of Wellington Faculty of Education Ethics Committee

Appendix 4 Ethics amendment



FACULTY OF EDUCATION
PO Box 17-310 Wellington 6012, New Zealand
Website www.victoria.ac.nz

9 August 2013

Anne Yates
Lecturer
Victoria University of Wellington Faculty of Education
C/- School of Educational Psychology and Pedagogy
Donald Street
Wellington

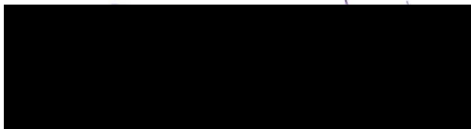
Dear Anne

RE: Ethics application SEPI/2013/60 Addendum to SEPI/2013/10 RM 19669

I am pleased to advise you that your ethics application addendum to '**An investigation of the relationships between teachers' beliefs and practices about assessment and students' achievement in NCEA.**', with the required changes, has been approved by the Victoria University of Wellington Faculty of Education Ethics Committee. Please note that the approval for your research to commence is from the date of this letter.

Best wishes for your research.

Yours Sincerely



Andrea Milligan

Co-Convener
Victoria University of Wellington Faculty of Education Ethics Committee

Appendix 5: Factor matrix for 4 factor solution

Factor Matrix^a

	Factor			
	1	2	3	4
Assessment provides information on how well schools are doing	.586			
Assessment places students into categories				
Assessment is a way to determine how much students have learned from teaching	.690			
Assessment provides feedback to students about their performance	.568		.302	
Assessment is integrated with teaching practice	.469			
Assessment results are trustworthy	.661			
Assessment forces teachers to teach in a way against their beliefs	-.369	.309	.513	
Teachers conduct assessments but make little use of the results		.306		
Assessment results should be treated cautiously because of measurement error		.319	.337	.305
Assessment is an accurate indicator of a school's quality	.526	.613		
Assessment is assigning a grade or level to student work				
Assessment establishes what students have learned	.696			
Assessment feeds back to students their learning needs	.467			
Assessment information modifies ongoing teaching of students	.466		.339	-.515
Assessment results are consistent	.644			
Assessment is unfair to students	-.512	.337		
Assessment results are filed & ignored	-.398			
Teachers should take into account the error and imprecision in all assessment				
Assessment is a good way to evaluate a school	.614	.638		
Assessment determines if students meet qualifications standards	.617			
Assessment measures students' higher order thinking skills	.713			
Assessment helps students improve their learning	.700			
Assessment allows different students to get different instruction	.327	-.344		
Assessment results can be depended on	.795			
Assessment interferes with teaching	-.474	.378	.435	
Assessment has little impact on teaching				
Assessment is an imprecise process	-.348	.328	.461	

Extraction Method: Maximum Likelihood.

a. 4 factors extracted. 5 iterations required.

Appendix 6: The anti-image correlation matrix for a four factor solution

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Assessment provides information on how well schools are doing	.853 ^a	-.239	-.112	-.029	-.192	-.079	.054	.023	-.084	-.161	-.143	.072	.232	-.046	-.101	-.066	-.071	.134	-.210	-.006	.056	-.207	.061	.017	.119	-.050	-.012
Assessment places students into categories	-.239	.598 ^a	.048	-.060	-.118	.096	-.075	-.006	-.003	.143	-.095	-.065	-.136	.047	-.077	-.026	-.117	.045	-.160	-.063	.099	.039	-.041	.041	-.080	-.037	.071
Assessment is a way to determine how much students have learned from teaching	-.112	.048	.860 ^a	-.195	-.050	-.195	-.128	-.094	-.110	.203	.057	-.312	.144	-.052	-.134	.152	.191	.106	-.182	.075	-.302	.013	-.063	.092	-.092	-.121	.017
Assessment provides feedback to students about their performance	-.029	-.060	-.195	.896 ^a	-.130	-.040	-.046	-.007	.112	.065	-.079	.009	-.213	-.316	.078	-.054	.030	-.074	.022	-.134	-.030	.116	-.051	-.103	.070	-.033	-.052
Assessment is integrated with teaching practice	-.192	-.118	-.050	-.130	.827 ^a	-.232	-.186	-5.397E-5	-.030	.093	.087	-.214	-.046	-.143	.022	-.075	.099	-.223	.033	.038	-.023	.136	-.116	.023	-.152	.137	.101
Assessment results are trustworthy	-.079	.096	-.195	-.040	-.232	.881 ^a	-.041	.138	.203	-.047	.026	.004	-.018	.199	-.090	-.007	-.043	.187	.061	-.127	.084	.002	.098	-.313	.104	.106	-.106
Assessment forces teachers to teach in a way against their beliefs	.054	-.075	-.128	-.046	.186	-.041	.829 ^a	-.085	-.179	.093	-.045	-.053	.080	-.108	.106	-.168	.103	-.007	-.053	-.014	.058	.130	-.069	-.073	-.299	.077	-.136
Teachers conduct assessments but make little use of the results	.023	-.006	-.094	-.007	-5.397E-5	.138	-.085	.692 ^a	.036	-.124	-.017	-.030	.081	.060	-.039	-.054	-.449	.220	.131	-.193	.175	.085	-.074	-.144	.018	-.004	-.287
Assessment results should be treated cautiously because of measurement error	-.084	-.003	-.110	.112	-.030	.203	-.179	.036	.684 ^a	-.187	-.101	-.016	.040	.165	.204	.025	.045	-.176	.181	-.116	.053	.010	-.049	-.214	-.061	.119	-.287
Assessment is an accurate indicator of a school's quality	-.161	.143	.203	.065	.093	-.047	.093	-.124	-.187	.659 ^a	.024	-.254	-.169	-.030	-.109	-.061	.161	.150	-.666	.164	-.178	.085	.009	.072	-.136	-.088	.089
Assessment is assigning a grade or level to student work	.143	-.095	.057	-.079	.087	.026	-.045	-.017	-.101	.024	.521 ^a	-.131	.052	-.159	-.021	.034	-.017	-.021	-.083	.043	-.203	-.073	.195	.056	-.184	-.071	.210
Assessment establishes what students have learned	.072	-.065	-.312	.009	-.214	.004	-.053	-.030	-.016	-.254	-.131	.894 ^a	-.042	.034	.124	.113	-.091	.065	.056	-.174	-.094	-.202	.009	.009	.102	-.116	-.030

Assessment feeds back to students their learning needs	.232	-.136	.144	-.213	-.046	-.018	.080	.081	.040	-.169	.052	-.042	.852 ^a	-.145	.029	.102	-.002	-.060	.035	.092	-.204	-.201	-.107	.026	-.027	-.030	-.016
Assessment information modifies ongoing teaching of students	-.046	-.047	-.052	-.316	-.143	.199	-.108	.060	.165	-.030	-.159	-.034	-.145	.829 ^a	-.189	.020	-.029	-.088	-.019	.083	.057	-.114	-.124	.018	.040	.192	.051
Assessment results are consistent	-.101	-.077	-.134	.078	.022	-.090	.106	-.039	.204	-.109	-.021	.124	.029	-.189	.863 ^a	.141	-.014	-.048	.137	-.106	-.043	.031	-.028	-.419	-.176	.146	-.099
Assessment is unfair to students	-.066	-.026	.152	-.054	-.075	-.007	-.168	-.054	.025	-.061	.034	.113	.102	.020	.141	.899 ^a	-.050	-.134	-.028	.112	-.189	-.014	.095	.089	-.220	.120	-.025
Assessment results are filed & ignored	-.071	-.117	.191	.030	.099	-.043	-.103	-.449	.045	.161	-.017	-.091	-.002	-.029	-.014	-.050	.772 ^a	-.214	-.119	.146	-.088	.096	.109	-.004	-.221	.030	.127
Teachers should take into account the error and imprecision in all assessment	.134	.045	.106	-.074	-.223	.187	-.007	.220	-.176	.150	-.021	.065	-.060	-.088	-.048	-.134	-.214	.496 ^a	-.135	-.117	.072	.020	.018	-.109	.169	-.066	-.282
Assessment is a good way to evaluate a school	-.210	-.160	-.182	.022	.033	.061	-.053	.131	.181	-.666	-.083	.056	.035	-.019	.137	-.028	-.119	-.135	.722 ^a	-.135	.097	.059	.096	-.247	.084	.049	-.115
Assessment determines if students meet qualifications standards	-.006	-.063	.075	-.134	.038	-.127	-.014	-.193	-.116	.164	.043	-.174	.092	.083	-.106	.112	.146	-.117	-.135	.883 ^a	-.239	-.134	.119	.000	.142	.006	.038
Assessment measures students' higher order thinking skills	.056	.099	-.302	-.030	-.023	.084	.058	.175	.053	-.178	-.203	-.094	-.204	.057	-.043	-.189	-.088	.072	.097	-.239	.883 ^a	-.049	-.006	-.161	.162	.068	-.120
Assessment helps students improve their learning	-.207	.039	.013	.116	.136	.002	.130	.085	.010	.085	-.073	-.202	-.201	-.114	.031	-.014	.096	.020	.059	-.134	-.049	.887 ^a	-.181	-.375	-.136	.087	-.064
Assessment allows different students to get different instruction	.061	-.041	-.063	-.051	-.116	.098	-.069	-.074	-.049	.009	.195	.009	-.107	-.124	-.028	.095	.109	.018	.096	.119	-.006	-.181	.859 ^a	-.069	.063	-.057	-.005
Assessment results can be depended on	.017	.041	.092	-.103	.023	-.313	-.073	-.144	-.214	.072	.056	.009	.026	.018	-.419	.089	-.004	-.109	-.247	.000	-.161	-.375	-.069	.854 ^a	.029	-.166	.241
Assessment interferes with teaching	.119	-.080	-.092	.070	-.152	.104	-.299	.018	-.061	-.136	-.184	.102	-.027	.040	-.176	-.220	-.221	.169	.084	.142	.162	-.136	.063	.029	.811 ^a	-.050	-.271
Assessment has little impact on teaching	-.050	-.037	-.121	-.033	.137	.106	.077	-.004	.119	-.088	-.071	-.116	-.030	.192	.146	.120	.030	-.066	.049	.006	.068	.087	-.057	-.166	-.050	.499 ^a	-.093
Assessment is an imprecise process	-.012	.071	.017	-.052	.101	-.106	-.136	-.287	-.287	.089	.210	-.030	-.016	.051	-.099	-.025	.127	-.282	-.115	.038	-.120	-.064	-.005	.241	-.271	-.093	.739 ^a

