PERCEPTIONS OF ACADEMIC LEADERS IN NEW ZEALAND REGARDING THE FUNCTIONS OF MODERATION OF INTERNAL ASSESSMENT:
A MIXED METHODS STUDY

BY

ANNA HUIA WILLIAMS

A thesis

submitted to the Victoria University of Wellington

in fulfilment of the requirements for the degree of

Doctor of Philosophy

Victoria University of Wellington

2019

Abstract

In formal education and training, internal assessment (in which assessor judgements are made within education organisations) is widely used for summative purposes to contribute to the award of qualifications. In many jurisdictions including New Zealand, organisations that conduct these high-stakes internal assessments are required by regulation to engage in moderation within the organisation and with external quality assurance bodies to quality-assure those assessments. However, policies are rarely implemented directly as intended. Instead, they are *enacted* by organisations, that is, policies are interpreted and translated, with multiple factors influencing this process. One such factor is the person who takes the role of 'policy narrator' and leads the policy interpretation and translation within the organisation. In New Zealand there is further potential for enactment variation because education organisations are largely self-governing, and thus have substantial freedom regarding organisational systems and practices. Moderation is commonly held to have both accountability and improvement purposes. However, it is unknown what policy narrators within New Zealand organisations consider the functions of moderation to be.

This study sought to explore what the academic leaders who are responsible for moderation in New Zealand secondary and tertiary organisations (i.e., those likely to be policy narrators) perceive as the functions of internal moderation and national moderation conducted by the New Zealand Qualifications Authority (NZQA moderation). Further, the study sought to ascertain whether there are any observable differences in perceptions according to organisation type. A pragmatic mixed methods sequential research design was implemented. An online survey instrument was developed informed from interview findings, and then administered using a census approach to collect data ($n = 221$). Both qualitative and quantitative data analyses were conducted.

Academic leaders were found to believe that moderation functions across multiple embedded contexts, from the immediate assessment event, to organisational and societal contexts. Internal and NZQA moderation were seen to work in the narrowly-focused area of assessment quality, and the broader areas of professional learning, organisational quality assurance, maintaining public and stakeholder confidence, and educational quality (internal moderation only). Instead of subscribing to the dominant improvement and accountability discourses, for the most part academic leaders tended think of moderation in more encompassing ways than the literature suggests.

Respondents from Private Training Enterprises (PTEs) tended to see the organisational quality assurance and educational quality functions as being more important or having a stronger emphasis, and to hold a broader view of moderation functions, than those from schools.

These findings could assist those in organisations to recognise and examine the influence of their own perceptions on practice, and identify opportunities to optimise how their organisations use moderation. The findings enable policy makers to ascertain the degree of alignment between policy intent and enactment, and could inform policy development and communication to the sector. Further, the potential for NZQA to increase the broader and improvement-focused aspects of moderation practice, while maintaining—and enhancing—its accountability focus is highlighted.

Acknowledgements

It is said that doing a PhD is an exercise in endurance and tenacity. I agree. It is. My time in the outdoors probably set me in good stead for the experience: I was used to hard work, drudgery, dealing with uncertainty, and keeping going when tired, disheartened, or not enjoying myself. I knew that there would be rewards: the fascination of exploration, gaining insights, new skills, and much deeper knowledge, and an immense sense of satisfaction and achievement once done.

As with any project that takes this length of time, a lot of other stuff occurs between starting out and submitting. In short, *life* happens. This research project has *been there* throughout. It has been a rich and mutable experience, variously engrossing, exhilarating, a trudge, progressing nicely, casting a long shadow, and offering welcome respite from the other things going on. I understand that this tapestry of experience is normal.

There are many people that I would like to thank for the part that they played in this journey. A huge thank you to:

- those who participated in the trial interviews, and piloted the survey instrument.
- the academic leaders who participated in this study.
- my supervisors, Dr Michael Johnston and Associate Professor Robin Averill, for your advice, feedback, patience, encouragement, challenging me to do better, and for our discussions. You have guided me through the quantitative (Michael), and qualitative (Robin) aspects. Together, you have helped me to navigate the process of bringing the different strands together, and have accompanied me through this project.
- Drs Sara Cotteral and Lizzie Towl in the Faculty of Graduate Research. The workshops that I attended provided me with strategies and insights that I have found invaluable throughout this journey.
- my whānau, friends, and work colleagues. Your support, encouragement, interest (or lack thereof), patience, distractions, laughter, and love helped keep me grounded, motivated, and refreshed when needed. A special thank you to Janet Stephenson for your feedback on my penultimate draft. It was affirming and a real confidence boost.
- our neighbourhood birdlife, especially the kākā, for the distraction and joy that they have provided me throughout this journey.
- Ian, for your unwavering and ongoing support, encouragement, patience, understanding, good humour, and love. You provide me with a base from which to work.

Contents

List of Figures

List of Tables

Glossary

Term (and acronym)	Definition
Academic leader	The person within each education organisation who is responsible for the management of internal moderation and the coordination of national moderation
Achievement standard	Assessment standard that is derived from, and aligned with, the NZ Curriculum or Te Marautanga o Aotearoa
Assessment	Evaluation or measurement of students' progress towards, or attainment of, intended learning outcomes
Assessment material	The instrument used to generate or collect student evidence, and all associated material that guides the assessment of that evidence and assessor judgements made
Assessment standard	Standard that defines specified knowledge, skills, or applications, and is set at the level on the New Zealand Qualifications Framework that best reflects those specifications
Assessor	See Teacher
Assessor judgement	The decision regarding the quality of student evidence as compared with the predefined criteria of the assessment standard, which determines the grade awarded
Course	Component part of a programme of learning, the predetermined outcomes of which are summatively assessed
Directory of Assessment Standards (DAS)	Register of nationally-set and quality assured assessment standards
Education	Refers to education and training
Education organisation or Organisation	Refers to Institutes of Technology or Polytechnics, Private Training Enterprises, and schools
External assessment	Assessment in which the assessor judgements regarding student evidence are made outside of an education organisation (e.g., by an external examiner)
External moderation	Moderation conducted between organisations
Industry Training Organisation (ITO)	A body corporate that acts as the standard-setting and quality assurance body for specific vocational areas, and arranges industry training; as defined by the Industry Training and Apprenticeships Act (1992)

Institute of Technology or Polytechnic (ITP)	As defined by the Education Act (1989); a publicly owned tertiary education organisation
Internal assessment	Assessment in which the assessor judgements regarding student evidence are made within an education organisation (e.g., by a teacher)
Internal moderation	Moderation conducted within an education organisation
Learning	Organisational or professional learning (i.e., on the part of teachers or the education organisation), as different from 'student learning' (see below)
Moderation	Quality assurance process that directly addresses the quality of internal assessment
National Certificate of Educational Achievement (NCEA)	New Zealand's national senior secondary qualifications
National moderation	Moderation conducted by quality assurance bodies against nation-wide or jurisdiction-wide assessment standards or syllabi
New Zealand Curriculum	The national curriculum for English-medium education in the primary and secondary education sectors
New Zealand Qualifications Authority (NZQA)	The government agency responsible for quality assuring non-university TEOs and the programmes they deliver, qualifications listed on the NZQF, internal assessment of achievement standards and NZQA-owned unit standards, and the internal assessment practice of schools, as well as administering the DAS and NZQF, among other things.
NZQA moderation	National moderation conducted by NZQA, mainly of achievement standards and NZQA-owned unit standards
New Zealand Qualification Framework (NZQF)	The national framework on which quality-assured secondary and tertiary qualifications in New Zealand are registered
Private Training Establishment (PTE)	As defined by the Education Act (1989); a privately-owned education organisation that operates in the tertiary sector
Quality assurance body	The body that is responsible for the quality assurance of certain education organisations and the services that they deliver (e.g., education, qualifications), or of internal assessment against certain groups of assessment standards

School	A registered State or State-integrated school, as defined by the Education Act (1989)
Standard-setting body	The body that is responsible for the development and maintenance of certain groups of assessment standards
Standards-based assessment	Assessment in which the quality of student evidence is compared with predefined criteria, and the grade awarded is based on that comparison
Student	Refers to student, learner, candidate, or trainee
Student evidence	Evidence generated through, or for, assessment by a student of their achievement or learning progress
Student learning	Learning and achievement of students, as different from organisational or professional learning (see 'learning' above)
Summative assessment	Assessment that is conducted to report on student achievement
Te Marautanga o Aotearoa	The national curriculum for Māori-medium education in the primary and secondary education sectors
Teacher	Refers to teacher, tutor, trainer, instructor, or assessor
Teaching	Refers to teaching, instruction, training, or delivery
Tertiary Education Organisation (TEO)	Education organisations operating in the tertiary (post-compulsory) education sector
Unit standard	Assessment standard that is not derived from, or aligned with, the NZ Curriculum or Te Marautanga o Aotearoa.

1. Introduction: Setting the scene

This thesis sets out to explore the perceptions of academic leaders in New Zealand education organisations regarding the functions of moderation of internal assessment, the quality assurance process that focuses on the assessment in which assessment judgements are made within the organisation. The academic leaders whose perceptions are explored are those with responsibility for and oversight of moderation within the organisation.[1] The education organisations involved are Institutes of Technology and Polytechnics (ITPs), Private Training Enterprises (PTEs), and state and state-integrated secondary schools (henceforth called schools).

1.1 Introduction to the research gap

In formal education and training, students are assessed to evaluate their learning progress and achievement. A core purpose of such summative assessment is to credential students: to ascertain whether they have acquired and demonstrated the intended knowledge, skills, and attitudes, to meet the specified requirements for a standard or course (Harlen, 2007). Internal assessment (in which assessment judgements are made within education organisations) is widely used internationally for summative assessment purposes, including to contribute towards qualifications (Crisp, 2013).

To be credible to stakeholders, qualifications, and the assessments that contribute towards them, must be robust and trustworthy (Broadfoot, 2007). Quality assurance processes, including moderation, are used to ensure an acceptable level of comparability between the assessments conducted by different organisations, and therefore, the qualifications awarded on the basis of those assessments. Moderation is a quality assurance process through which the quality and integrity of internal assessments are intended to be ensured (Harlen, 2007; Newton & Shaw, 2014). In many jurisdictions, including New Zealand, education organisations that conduct internal assessment for credentialing purposes[2] are required both to implement moderation within the organisations themselves, and to engage in moderation conducted by external quality assurance bodies.

---

[1] Leadership and management, as distinct fields of study in their own right, are outside of the scope of this thesis.

[2] Where credentialing purposes are taken to mean for summative assessment against nationally-set assessment standards, or awards and qualifications that are registered on the New Zealand Qualifications Framework. Refer 1.2.1 for details.

Previous research has found that moderation can fulfil quality control and accountability functions with respect to internal assessments, and that it assists the maintenance of public and stakeholder confidence in internal assessments, assessment standards, and associated qualifications (see 2.4.3). Previous research has also found that moderation can have improvement and learning functions, such as providing opportunities for professional development about assessment (see 2.4.4).

Scholarship in the policy implementation field has established that policies (such as the requirement for education organisations to engage in moderation) are rarely implemented exactly as intended by policy makers. Instead, organisations develop their own understandings of policies, and enact them accordingly. The ways in which organisations understand and enact policies are influenced by multiple factors, which can lead to considerable variation in the resulting enactment (Braun, Maguire, & Ball, 2010). For each organisation, these influencing factors include the context (Braun, Ball, Maguire, & Hoskins, 2011), and the people involved in the policy enactment. The role that different people take in relation to policy work within an organisation varies in part with people's positions of responsibility (Ball, Maguire, Braun, & Hoskins, 2011a). Those involved in interpreting policy messages and determining an organisation's response do so through their own existing knowledge and understanding (Spillane, Reiser, & Reimer, 2002). Thus, what those in the active policy work roles within organisations know and understand influences the organisation's understanding and subsequent enactment of policies.

1.1.1 The research gap

Little research has been conducted into what academic staff within organisations understand about moderation and its functions. Recent qualitative studies by Adie, Lloyd, and Beutel (2013), and Grainger, Adie, and Weir (2016) explored the perceptions of academic staff in the Australian university sector of internal moderation (see 2.4.6 for details). Beyond these two studies, no research appears to have investigated the perceptions of academic staff or leaders into the functions of moderation, let alone in the New Zealand context. The present thesis aims to address this research gap by exploring perceptions about the functions of moderation, of academic leaders in certain education organisations in New Zealand. Because people's positions of responsibility impact on their policy work (Ball et al., 2011a), the population whose perceptions are explored in the present study comprises the academic leaders with responsibility for the management of

internal moderation and the coordination of national moderation within organisations.[3] These academic leaders are arguably the most likely to have organisation-wide views of internal moderation and of the national moderation that the organisation engages with, as well as an understanding of the system-level, structural requirements that exist. The findings of Ball et al. (2011a) also suggest that these academic leaders are likely to possess a greater level of power, agency, influence, and control over the moderation processes and practices enacted within their organisations than those without responsibility for moderation.

The rest of this chapter sets the scene for the study by providing an overview of the New Zealand education and national qualifications systems (the context of the study), before introducing a theoretical construction to assist conceptualising that context. The chapter then outlines the journey of this researcher to the present study, before providing an overview of the structure of this thesis.

1.2 Setting the scene: Context

1.2.1 The New Zealand education and national qualifications systems

New Zealand is a comparatively sparsely populated nation of 4.84 million people with a land area of 271,000 square kilometres in the South Pacific (Statistics New Zealand, n.d., 2018). The centres of population (and education organisations of interest to the present study) are geographically dispersed across the length and breadth of the country. The New Zealand education system comprises early childhood education, primary (Years 1–8), and secondary (Years 9–13) school sectors,[4] and a tertiary (also called 'post-compulsory', or 'further education') sector. The national curriculum that applies across the primary and secondary sectors is called the New Zealand Curriculum.[5] As per the Education Act (1989) and its subsequent amendments, the policy framework and strategic direction of the New Zealand

---

[3] In different organisations, this responsibility will be incorporated within different roles: The Principal's Nominee in schools, who may be a Deputy or Assistant Principal, Dean, or other member of the senior management team; in the tertiary sector this responsibility is commonly held by an Academic Manager or Director, or Quality Manager.

[4] The primary and secondary sectors are also called the 'compulsory sector' as schooling is compulsory in New Zealand between the ages of six and 16 (Ministry of Education, 2018a).

[5] The New Zealand Curriculum applies for English-medium schooling, and Te Marautanga o Aotearoa applies for Māori-medium schooling (Ministry of Education, 2018a).

education system is set by central government and administered by Crown agencies and entities including the Ministry of Education, the Education Review Office, the New Zealand Qualifications Authority (NZQA), and the Tertiary Education Commission (Ministry of Education, 2015).

There is a range of education organisation types in the New Zealand secondary and tertiary sectors. In the secondary sector, these include state and state-integrated schools (publicly-owned and funded),[6] private schools (privately-owned, but receive some public funding), and kura kaupapa Māori, which provide Māori-medium education and are publicly-owned and funded (Ministry of Education, 2018b). Tertiary education organisations (TEOs) include ITPs, universities, and wānanga (all publicly-owned and partly-funded), registered and non-registered PTEs (privately-owned; if registered, may receive some public funding), and Industry Training Organisations (ITOs), which represent particular industries, and are industry- and publicly-funded (Industry Training and Apprenticeships Act, 1992; Ministry of Education, 2015).

The New Zealand Qualifications Framework (NZQF) is the framework on which quality-assured secondary and tertiary qualifications in New Zealand are registered. It is legislated in the Education Act (1989) and is administered by NZQA (NZQA, 2016b). The framework focuses on outcomes (defined knowledge and skills, and their application), and comprises 10 levels. Each consecutive level describes increasingly complex and demanding knowledge, skills and applications. Naming conventions, credit and level requirements, and other design aspects are stipulated for all qualifications that are registered on the NZQF. A qualification is listed on the NZQF at the level that best aligns with its graduate profile, which defines what a graduate with the qualification should know and be able to do. These features are intended to enable clear communication of what a graduate of a qualification has demonstrated by way of achievement and performance (NZQA, 2016b). At the time of data collection for the present study, both local (i.e., organisation-specific) and national qualifications were registered on the NZQF.

Nationally-set and quality-assured assessment standards that are registered on the Directory of Assessment Standards (DAS) can contribute towards national qualifications (NZQA, n.d.g). Assessment standards define specified knowledge, skills, or applications that a student who has been awarded the standard has demonstrated. Each assessment

---

[6] Across the compulsory sector, 96% of students attend these types of schools (Ministry of Education, 2015).

standard is set at the level of the NZQF that best reflects the complexity and demand of its assessment outcomes and carries a defined number of 'credits' indicating the nominal learning hours associated with that standard. There are two types of assessment standard: achievement standards and unit standards. Achievement standards are derived from, and aligned with, the New Zealand Curriculum,[7] and can be achieved at three grades (Achieved, Merit, Excellence). The Ministry of Education is the standard-setting body for achievement standards. Unit standards are not curriculum-aligned,[8] and while some can be achieved at the same three grades as achievement standards, Achieved is the only grade available for a large majority. Various standard-setting bodies develop, maintain, and quality-assure the assessment of different groups of unit standards. For example, NZQA is the standard-setting body for the English Language unit standards, and the Building and Construction Industry Training Organisation is the standard-setting body for the Construction unit standards. Once a student has achieved an assessment standard, credits for it are recorded on that student's centrally held Record of Achievement.

The national senior secondary qualifications are the National Certificates of Educational Achievement (NCEAs), against which most secondary schools assess students (NZQA, n.d.b). There are three levels of NCEA certificate (Levels 1, 2 and 3), each of which can be awarded based on externally assessed or internally assessed standards (see below), or a combination of both. The NCEAs can be gained through a broad range of subjects and assessment standards. The requirements for NCEA at each NZQF level comprise a total number of credits (of which a minimum number must be at that level) and certain literacy and numeracy requirements. Beyond these parameters, there are no limits on the assessment standards that can be used to meet the balance of credits required. Provision is made for higher levels of achievement to be recognised, via an endorsement (Merit or Excellence) at certificate and course levels. Most students are assessed for NCEA during their last years at secondary school, although some continue to gain credits towards NCEA once in tertiary study (NZQA, n.d.b). By itself, NCEA Level 3 does not satisfy the entry requirements for New Zealand university study. Instead, the University Entrance award is needed. To meet the requirements of University Entrance, students need to have NCEA Level 3, plus attain at

---

[7] For the English-medium standards. The Marautanga achievement standards are set against Te Marautanga o Aotearoa (Ministry of Education, 2015).
[8] Previously, some unit standards were aligned with the New Zealand Curriculum. However, the 2008 review of curriculum-derived standards resulted in expiration of all curriculum-derived unit standards, and the review and development of new achievement standards. The resulting Level 1 achievement standards were introduced in 2011, and the Level 2 and Level 3 achievement standards in 2012 and 2013 respectively (NZQA, 2012c).

least 14 credits in each of three approved subjects, and meet specified literacy and numeracy requirements (NZQA, n.d.g).

TEOs (including ITPs and PTEs) generally focus on delivering vocational and technical education and training, and there tends to be little delineation between the types of provision each TEO type offers (Ministry of Education, 2015). TEOs offer a wide variety of local and national awards and qualifications, across a diverse range of disciplines, and across most levels of the NZQF (from transitional or 'bridging' education, to post-graduate level programmes). The offerings are comprised of local courses, achievement or unit standards, or other nationally—or internationally—set syllabi or prescriptions. Thus, there is immense diversity in the education provision among TEOs in New Zealand.

1.2.2 Quality assurance of the education and national qualifications systems

Quality assurance systems are in place across all levels of the education system, and implemented by the agencies, entities, and quality assurance bodies responsible (Ministry of Education, 2015). The two main quality assurance bodies[9] for the organisations involved in the present study are the Education Review Office[10] and NZQA,[11] of which NZQA is the main interest to this study. NZQA is responsible for quality-assuring TEOs[12] and the programmes they deliver, qualifications listed on the NZQF (see 2.3.1), internal assessment of NZQA- and Ministry-of-Education- owned and managed assessment standards (via moderation, see 2.4), and the internal assessment practices of secondary schools (see 2.3.1). NZQA is also responsible for administering the NZQF, DAS, and Records of Achievement (see 1.2.1), managing the external assessment of achievement standards, and being the standard-setting body for NZQA-owned unit standards (NZQA, n.d.d).

---

[9] Although ITOs are responsible for the quality assurance of their unit standards.
[10] The Education Review Office is responsible for quality-assuring the education and care provided to students in organisations in the early-childhood, primary, and secondary sectors (Education Review Office, 2014).
[11] Although historically, the quality assurance body for ITPs was Institutes of Technology and Polytechnics Quality. Refer 2.3.1.2.
[12] Non-university TEOs only. Universities are quality assured by Universities New Zealand (Ministry of Education, 2015).

1.2.3 Internal assessment and its quality assurance: Moderation

Internal assessment is assessment that has been conducted and judged by assessors internal to an education organisation, for instance, teachers (Harlen, 2007). In some cases, the assessment activities have also been designed and developed within an organisation, whereas in other cases these are purchased or obtained from another source.[13]

Moderation is the quality assurance process that directly addresses the quality of internal assessment. Through moderation, the quality and integrity of assessments and the fairness and robustness of assessment processes used are ensured, thus enabling the credibility of the associated assessment standards and resulting qualifications to be upheld (Adie et al., 2013; Crisp, 2017; Crooks, 2011). In the senior secondary and tertiary sectors of the New Zealand education system, organisations are required to implement internal moderation (within the organisation) as part of their quality management system, although how that internal moderation is to be conducted is not prescribed. Those assessing against nationally-set assessment standards are also required to engage in national moderation with relevant standard-setting or quality assurance bodies (NZQA, 2011b, 2011c, 2015, 2017a). Of interest to the present study are internal moderation, and the national moderation as conducted by NZQA (henceforth called NZQA moderation).

1.2.4 Self-managing education organisations

As a result of the 1988 policy initiative *Tomorrow's Schools,* the New Zealand education system is characterised by a devolved and decentralised approach to governance and management,[14] whereby organisations in all sectors are self-managing[15] (Ministry of Education, 2015; Openshaw, 2014). This approach has implications for the ways in which each organisation enacts policies, including those pertaining to quality assurance and moderation. Under *Tomorrow's Schools*, schools are governed by Boards of Trustees (comprised mainly of elected parents) and have flexibility in how they apply the New Zealand

---

[13] In New Zealand, assessment materials for assessment standards delivered extensively by schools are provided free-of-charge by the Ministry of Education (via the Te Kete Ipurangi—TKI—website) and NZQA. These assessment materials are widely used in the school sector.

[14] Although the control exerted by the Ministry of Education (over funding and staffing) and the Education Review Office and NZQA as quality assurance bodies (in the criteria applied with evaluating a school's performance) is considered by some (e.g., Slowley, 2008) to indicate a level of centralised control.

[15] This may be about to change. The recommendations for reform of the education sector that were recently made by an independent taskforce (and currently under consultation) include shifting away from schools being self-governing, towards a more centralised model (Redmond, 2018).

Curriculum to best meet the needs of their learners and communities. Publicly owned TEOs (including ITPs and universities) are autonomous and are governed by councils comprising local or central government representatives, and members of the community, business, staff, and students. PTEs are, by definition, private enterprises, and are self-governing and self-managing. The way in which each organisation enacts policies is shaped by its individual context, including size, location, funding, staffing, student cohort, local community, existing conditions and commitments, and external pressures, expectations, and support (Braun et al., 2011). It follows that the New Zealand situation of self-managing organisations has likely resulted in greater levels of difference and nuance in policy enactment between organisations, than would be the case in a less devolved governance and management situation. Thus, the result is likely to be multiple self-managing organisations, enacting policies and practices in ways ranging from relatively standardised to unique to that organisation. Of relevance to the present study is the enactment of moderation policies: While there may be similarities in some of the internal moderation practices and approaches used within organisations, there is unlikely to be a uniform approach, procedure, or enactment across the organisations of interest.

1.2.5 Layers of embedded contexts

The scene in which the present study is set (described in 1.2.1–1.2.4) can be conceptualised as an overarching context containing layers of embedded contexts. The overarching context in this conceptualisation is New Zealand society. Within this, the education and national qualifications systems form the first embedded context, with the NZQF, DAS, and NCEAs that comprise these systems (1.2.1), along with the quality assurance bodies (1.2.2) and associated regimes (including national moderation; 1.2.3), all being part of this layer. Embedded within the education and national qualifications systems context, individual education organisations form another contextual layer. The different types of self-managing education organisations within New Zealand (1.2.1, 1.2.4), mean that diversity and heterogeneity across those organisations in quality management processes (e.g., internal moderation; 1.2.3) are likely. Embedded within the contexts of individual organisations are the education and programmes of learning that individual students experience, and it is within this micro-context that internal assessment occurs. The present study focuses on moderation of internal assessment, and ways in which those assessments compare across multiple organisations in the education system (see 1.1). Thus, multiple layers of embedded contexts (as described above) are relevant to the present study.

1.3 Setting the scene of this thesis

1.3.1 This researcher's journey to the present study

I bring insider and outsider perspectives (Bogdan & Biklen, 2007) to this study. In the mid-2000s I was teaching adult education in the context of outdoor education in an ITP. As a tutor I developed courses and assessments, taught and assessed courses, and engaged with internal and national moderation. The ITP primarily delivered vocational education and training in various disciplines, against local courses, unit standards, and other national and international prescriptions (e.g., New Zealand Diploma in Business prescriptions, NZQA, 2004; Food Preparation and Cooking qualifications, City & Guilds, n.d.). Prompted in part by a series of non-compliant national moderation results, as well as an upcoming Institutes of Technology and Polytechnics Quality academic audit, the ITP increased focus on quality assurance and implemented a suite of quality control and improvement initiatives. I became the Assessment and Moderation Coordinator for the organisation, coordinating and conducting internal moderation across the ITP, coordinating national moderation submissions to the quality assurance bodies and ITOs that the ITP dealt with, and training staff in assessment and moderation. Through the increased focus on quality assurance matters, emphasis was placed on control, improvement, and learning. We used moderation to assist this endeavour, including as a lens through which to identify aspects of courses, delivery, and assessment for improvement, and where professional development or support was required. Internal moderation, the associated conversations between moderators and tutors, and moderators and tutors working together to address identified issues, acted as avenues through which to provide professional support and development for tutors. We used national moderation results and feedback to assist improvement where possible.

From 2010 through 2014, I was employed by NZQA as an Assessment and Moderation Facilitator, firstly in the Tertiary Assessment and Moderation business unit,[16] and then in the Assessment and Moderation Services business unit.[17] My role included managing the national moderation systems for three sets of newly-developed unit standards (Literacy, Numeracy, and English for Academic Purposes), which were assessed in both the secondary and tertiary sectors. These standards were of high political interest because they

---

[16] The role of Tertiary Assessment and Moderation included monitoring the NZQA moderation results of TEOs, and moderating New Zealand Diploma in Business prescriptions (assessed by TEOs) and certain groups of NZQA-owned unit standards (some of which were assessed only by TEOs, and some by both TEOs and schools).

[17] The role of Assessment and Moderation Services included moderating achievement standards and some groups of NZQA-owned unit standards. Most assessment standards moderated by Assessment and Moderation Services were assessed by schools.

were used to assess the literacy and numeracy requirements for NCEA Level 1[18] and University Entrance.[19] I was responsible for tailoring the moderation approaches, training moderators, and managing NZQA moderation, for these sets of standards. I was also closely involved in developing interpretation and guidance resources for the standards, and in developing and delivering best practice workshops to introduce the standards to the secondary and tertiary sectors, and support and enhance their assessment.

Through my involvement with moderation in an ITP (an 'insider' perspective; Bogdan & Biklen, 2007) and then at NZQA (an 'outsider' perspective), I became increasingly aware that people in organisations seemed to see moderation in different ways, and to use it for different functions. From my own experience in an ITP, I knew that the uses to which we put internal moderation were often different from the uses to which we were able to put national moderation results and feedback. I also had observed that tutors and assessors often seemed to have limited agency in the moderation process; instead, moderation was 'done to them'. They also tended to focus on their own courses and programmes, with limited awareness of organisational contexts.

At NZQA, I was inducted into, and worked within, the culture of national moderation in two different business units, and with the implicit and explicit directives given to moderators regarding aspects of submissions that they could attend to and comment on. For example, a pronounced difference in the moderation conducted by the two business units concerned the ways in which the assessment materials in submissions were considered. In the years preceding and including the data collection period for the present study, the approach taken by Tertiary Assessment and Moderation was to moderate the assessment materials, and to provide a moderation result and feedback pertaining to those materials (NZQA, 2014c). In contrast, Assessment and Moderation Services moved away from that approach: From 2012, assessment materials in a submission were only moderated if it appeared that an issue with the assessor judgements or grades awarded was due to an issue with the assessment materials. Otherwise, the materials were not moderated, and no feedback for

---

[18] From 2012, the literacy and numeracy requirements for NCEA Level 1 could be met either by 10 credits from specified achievement standards for each of literacy and numeracy, or by the newly introduced suites of three literacy and three numeracy unit standards. The preceding year was a transition year, during which the literacy and numeracy requirements for NCEA level 1 could be met through the previous or new requirements (NZQA, n.d.c). From 2013 and 2014 respectively, the NCEA Level 1 literacy and numeracy requirements needed to be met in order to gain NCEA Levels 2 and 3 (NZQA, 2012b).

[19] As well as via certain achievement standards, the literacy requirements for University Entrance (of 5 credits in each of reading and writing at Level 2 or above) could be met via specified English for Academic Purposes unit standards, and the numeracy requirements of 10 credits could be met via the three Level 1 numeracy unit standards (NZQA, n.d.h, 2018).

them was given—but a compliant result was awarded (NZQA, 2011a). This approach was taken further in 2014, when Assessment and Moderation Services began using a modified moderation report form from which the assessment materials section was omitted: The materials were not moderated and no moderation result or feedback regarding them was given, unless an issue with judgements or grades awarded appeared to be due to an issue with those materials (NZQA, 2014e).

Observing the nuances in moderation approach and feedback within, and between, the NZQA business units, I wondered what people in client-organisations made of the differences. Through conversations in passing with various attendees at NZQA workshops that I delivered, I sensed that moderation was used for multiple and varying functions in organisations. I realised that in NZQA, we did not know what those in organisations considered the functions of either internal moderation or NZQA moderation to be. When I began this research study, I was still in the employ of NZQA, and set out to explore these issues.[20] I did so using a pragmatic mixed-methods sequential design, in which semi-structured interviews were conducted and the findings used to inform the development of a survey instrument. The online survey was administered to a sample of academic leaders, and qualitative and quantitative analyses conducted on the collected data.

1.3.2 Structure of thesis

The remainder of this thesis comprises five chapters. In Chapter Two, a review of recent literature relating to assessment, quality assurance in education, moderation, and policy implementation is presented. The chapter culminates with consideration of why we would want to know what academic leaders perceive the functions of moderation to be. Chapter Three presents the methodological foundations and the research methods of the present study. The results are presented in Chapter Four: Firstly, the results of quantitative analyses are presented, followed by the results from qualitative analyses (which are presented thematically). Within each theme, the results regarding moderation-in-general, internal moderation, and NZQA moderation are addressed, and reference is made to relevant quantitative results and literature, as part of focused discussions about the meaning of those results. In the penultimate chapter, the present study's research questions (introduced in 2.6.1) are revisited and answered using this study's results. In the final chapter, implications and conclusions relating to the present study's findings are presented.

---

[20] I subsequently left NZQA to pursue my studies.

2. Literature review

This chapter reviews relevant literature to distil the research gap that the present study sets out to address; that is, what academic leaders in New Zealand perceive the functions of moderation to be. Initially, assessment in formal education is introduced, with a focus on summative and standards-based assessment. Desirable properties of assessment are identified and considered, as are assessment methods and conditions. Internal assessment is the focus of the second section: Attention is paid to the process through which assessor judgements are made, and to the reasons why internal assessment is conducted by education organisations. Quality assurance in education is traversed in the third section, and the quality assurance processes that indirectly address internal assessment are discussed. The subsequent section focuses on moderation of internal assessment, with different approaches to moderation, moderation within and between organisations, and moderation at a jurisdiction-level examined. The use of moderation for accountability and improvement purposes are investigated, and the question of whether moderation can genuinely have both purposes is considered. Research findings suggesting that moderation is sometimes viewed with more nuance than simple accountability and improvement functions is then examined. The penultimate section explores the enactment of policy within education organisations. The final section articulates the research gap that has been identified through this literature review, before explicating the research questions that the present study seeks to address.

2.1 Assessment in formal education

A fundamental purpose of formal education is to equip students with knowledge, skills, and attitudes to assist them as they journey through life (Kemmis, Cole, & Suggett, 1998; Morshead, 1995). As they progress through formal education, students engage in a sequence of programmes of learning, designed to assist their learning journey, and help them to develop the requisite knowledge, skills, and attitudes to successfully participate in society and in their chosen vocations (European Association for Quality Assurance in Higher Education et al., 2015; Misko, Halliday-Wynes, Stanwick, & Gemici, 2014; Morshead, 1995). Programmes of learning usually comprise one or more component parts, called, for example, 'courses', 'modules', or 'papers'.[21] Generally, courses focus on coherent bodies of

---

[21] Throughout this thesis, programmes of learning are referred to as 'programmes' or 'programmes of learning', and their component parts are referred to as 'courses'.

knowledge and skills. Often they have predetermined specifications that define the intended learning outcomes and content, and how students will be assessed (Misko, 2015a).

A core purpose of teaching is to assist students to achieve the intended learning outcomes, and the basic functions of assessment are to evaluate students' progress towards, and attainment of, those outcomes (Miller, Linn, & Gronlund, 2009). The taught curriculum should reflect the specified curriculum and content. The pedagogical approach and activities should assist students to meet the intended learning outcomes and be successful in their learning endeavours. Assessment should fairly and credibly determine the extent to which students have achieved the intended learning outcomes (Biggs & Tang, 2007). Thus, teaching, student learning, and assessment are interrelated in the context of formal education. Assessment occurs within the context of the delivery of courses and programmes of learning, to assist and evaluate student learning (Harlen, 2007; Miller et al., 2009).

At secondary and tertiary levels of education, programmes of learning often lead to specific qualifications. As well as denoting the level and area of achievement, a qualification might also act as a 'stepping stone' to further study or employment (Misko, 2015a). Thus, stakeholders with an interest in the quality and outcomes of education often include not only those directly involved (e.g., students and their families, and the teachers and other staff in the organisations), but also the destination education organisations, employers, and industry bodies (European Association for Quality Assurance in Higher Education et al., 2015; Misko, 2015a).

2.1.1 Summative assessment

Assessment involves collecting evidence of student achievement (henceforth referred to as *student evidence*), and inferring meaning from that evidence by evaluating it against criteria or norms (Broadfoot, 2007; Harlen, 2007; Klenowski & Wyatt-Smith, 2014; Miller et al., 2009). Assessment is conducted for two main purposes: to inform teaching and learning (for formative purposes), and to report on student achievement (for summative purposes; Harlen, 2007). Formative assessment is integrally linked to teaching and student learning, and generally occurs throughout courses. It is used to assist students' learning by providing feedback, which is used by teachers and students to inform learning and teaching (Biggs & Tang, 2007; Broadfoot, 2007; Crooks, 2011; Harlen, 2007; Miller et al., 2009). Summative assessment is usually associated with awarding grades, and generally occurs at the end of blocks of student learning within courses (Biggs & Tang, 2007; Broadfoot, 2007; Crooks,

2011; Harlen, 2007; Miller et al., 2009). Summative assessment is often considered to be 'high stakes' due to consequences associated with it and, as such, is generally subjected to some form of quality assurance, (e.g., moderation; Broadfoot, 2007; Harlen, 2007). Multiple uses are made of the information provided by summative assessment, including the certification or credentialing of students' learning, and accountability (e.g., of teachers or organisations) for student achievement (Broadfoot, 2007; Crooks, 2011; Harlen, 2007; Miller et al., 2009).

Moderation is usually concerned with assessment that has a summative purpose (Broadfoot, 2007; Harlen, 2007). However, summative assessment is never divorced from student learning; it impacts on their learning, and all assessment is, therefore, at least implicitly formative because it has a developmental impact on students (Broadfoot, 2007; Crooks, 2011; Miller et al., 2009). For example, the results of summative assessments can impact on students' options for future study or employment, or on their self-concepts and self-confidence as learners. Furthermore, assessment for formative and summative purposes are sometimes intertwined; assessment events within courses are sometimes used to serve both purposes (Crooks, 2011). Thus, moderation can impact on the formative functions of assessment.

2.1.2 Standards-based assessment

Assessor judgements are made by evaluating student evidence against a norm or a criterion (Crisp, 2013; Harlen, 2007). When assessment evidence from each student is compared with other students in a class, a wider cohort, or a norming group, the assessment is 'norm-referenced'. When compared with specified criteria, the assessment is 'standards-based' or 'criterion-referenced' (Harlen, 2007; Miller et al., 2009).

In standards-based assessment, the quality of student evidence is compared with predefined criteria. The grade awarded is based on that comparison, and is supposed to be independent[22] of the quality of other students' work (Harlen, 2007; Miller et al., 2009). Two commonly-used forms of standards-based assessment are competency-based assessment (under which student evidence is judged against a criterion that defines only one level of quality), and achievement-based assessment – under which evidence is judged against criteria that define multiple levels of quality or proficiency (McMillan, 2007).

---

[22] Some research has suggested that comparison with other students' performance plays a role in assessor judgements in standards-based assessment. Refer 2.2.1.

2.1.3 Properties of good assessment

Scholars agree that certain properties characterise sound and trustworthy assessment. While the definitions and terminology used to describe the properties of assessment (e.g., validity, reliability, fairness) are contested, there is general agreement about what those properties, when taken together, comprise (Biggs & Tang, 2007; Newton & Shaw, 2014). All are considerations for moderation processes.

Validity is a cardinal concept in assessment, and is widely accepted as fundamental to evaluating assessment quality (Broadfoot, 2007; Cizek, 2016; Harlen, 2007; Miller et al., 2009; Newton & Shaw, 2014). However, there are substantially divergent views about the meaning of the term 'validity', and the debate is long-standing and ongoing (Cizek, 2016; Lissitz & Samuelsen, 2007; Newton & Shaw, 2014, 2016). Newton and Shaw (2014, 2016) identified a series of definitions, from one that is very narrow, with each subsequent definition more encompassing and broader than the definition prior to it in the sequence. The narrowest, subscribed to by Harlen (2007), relates only to an assessment instrument and the degree to which it enables the collection of evidence pertaining to what it was intended to assess. Under a slightly broader definition, validity is the degree to which an assessment measures what it is intended to measure: the *assessment outcome* (i.e., the assessor's judgements), as well as the assessment instrument and collection of evidence (e.g., Broadfoot, 2007). A still broader definition of validity, as subscribed to by Black, Harrison, Hodgen, Marshall, and Serret, (2010, 2011), pertains to the use made of assessment results, as well as to the results themselves (and the interpretation of what the evidence showed, and inferences about the meaning of those results). Under this perspective, it is the *decisions* made based on results that must be validated (Newton & Shaw, 2016). The broadest definition identified by Newton and Shaw (2016) conceptualises validity as relating to the *consequences of decisions made* based on the assessment results, as well as the interpretation of those results, and the uses made of them (e.g., Biggs & Tang, 2007; Crooks, Kane, & Cohen, 1996; McMillan, 2007; Miller et al., 2009).

Reliability is another important property of assessment. Reliability concerns the consistency of measurement: over time, between assessors, and between equivalent assessment instruments (Biggs & Tang, 2007; Broadfoot, 2007; Harlen, 2007; Miller et al., 2009; Newton & Shaw, 2014). Technically, reliability relates to the measurement error that is present in an assessment process, (i.e., the component of an assessment result that does not reflect the student's "*true* or *real* ability or skill" in the assessed domain; McMillan, 2007, p. 71, emphasis in original). A small measurement error value indicates that the reliability of the

assessment process is high. Measurement error can come from sources that are internal or external to the assessed student. Internal sources include the student's health and motivation levels. External sources include the clarity of assessment instructions, the types of response required by the assessment, assessor differences, and assessor biases (McMillan, 2007). Approaches to the evaluation of reliability include estimates of the consistency of assessor judgements, of the stability of assessment results over time or between equivalent forms of assessment, and evaluations of whether all items in an assessment are likely to measure the same trait (Biggs & Tang, 2007; McMillan, 2007; Miller et al., 2009). In internal standards-based assessment, reliability is primarily concerned with the consistency of assessor judgements: over time, across different sets of evidence, and between assessors (Biggs & Tang, 2007).

The relationship between the concepts of validity and reliability is debated. One school of thought considers reliability to be a subsidiary concept of validity, where reliability is one facet of validity, and is necessary to obtain fully valid assessment results. This aligns with the broader definitions of validity as discussed above: A lack of reliability would limit the extent to which decisions and inferences could be made, and may result in inappropriate consequences. Miller et al. (2009) took this position. A second school of thought holds the concepts as being separate. This aligns with the narrowest definition of validity – that validity only pertains to whether an assessment instrument collects the intended evidence, and does not consider reliability, precision, or bias (Newton & Shaw, 2016). According to this second view, it is possible to have assessments that are valid and biased, or valid and unreliable. Harlen subscribed to this perspective and extrapolated it, arguing that validity is separate from, but interdependent with, reliability. She saw a direct tension between the two: "For all assessment where both reliability and validity are required to be optimal, there is a trade-off between these two properties" (2007, p. 54). Thus, this second perspective holds that assessments with high validity will tend to have lower reliability, and vice versa: Assessment evidence that appropriately represents domains including application of knowledge is challenging to judge consistently, where evidence that can easily be consistently judged is of a more limited range of domains, for example, factual knowledge. The concept of 'dependability', as "the extent to which reliability is optimised while ensuring validity" (Harlen, 2007, p. 24) coheres with this perspective.

Scholars including Allal (2013), Brookhart (2003), and Moss, Girard, and Haniford (2006), have argued that psychometric definitions of validity and reliability, (i.e., those of the educational and psychological measurement field), are not fit for purpose when applied to internal assessment. However, some of their arguments appear only to concern methods for

evaluating the validity or reliability of assessments in large-scale standardised (norm-referenced) assessment schemes. Different methods are used for evaluating internal standards-based assessment (as per the context of this study), such as estimating inter-rater reliability, as opposed to test-retest, equivalent forms, or split-half methods (Miller et al., 2009). According to Biggs and Tang, the fundamental questions at the heart of evaluating reliability and validity are the same, irrespective of whether an assessment is norm-referenced or standards-based, large or small in scale: "Can we rely on the assessment results—are they reliable? Are they assessing what they should be assessing—are they valid?" (2007, p. 188). Furthermore, Brookhart (2003), and Moss et al. (2006), characterised internal assessment as primarily having a formative purpose, i.e., to improve teaching and student learning. As Harlen (2007) noted, the concept of reliability is not particularly important for formative assessment; teachers have opportunities to collect further evidence and give further feedback to students to correct misjudgements. However, the focus in the present study is summative assessment, to which Harlen's point seems not to apply. Nonetheless, what can be taken from these challenges is that not all approaches used to evaluate validity and reliability of large-scale norm-referenced tests will be appropriate or feasible for use with small-scale, standards-based assessments that involve a variety of assessment methods.

Fairness is another property of assessment that is considered in the literature. A fair assessment is unbiased and non-discriminatory in the assessment content, procedure, and judgement process. It has no interference from factors extraneous to the criteria being assessed (McMillan, 2007; Miller et al., 2009). The opportunities that students have to learn the assessible content is also considered to be an aspect of fair assessment. Rasooli, Zandi, and DeLuca, in their systematic meta-ethnography of literature, reported that one interpretation is that all students must have *equal* (i.e., the same) opportunities to learn, where another is that students must have *equitable* opportunities to learn (i.e., "differentiated instruction based on need" to ensure that all students are supported to succeed in learning; Rasooli et al., 2018, p. 170). For some scholars (e.g., Miller et al., 2009), fairness and issues of bias are encompassed in the concept of validity, while for others (e.g., Lissitz & Samuelsen, 2007), these factors pertain to reliability.

Determining the degree to which an individual assessment, or assessment system, engenders the properties discussed above is central to evaluating its quality. Crooks et al. (1996) depicted assessment as a chain comprising multiple links, from the design of an assessment instrument, to the impact of decisions made based on the results. This analogy is instructive because it implies that threats to the quality of assessment can occur with any

link in the chain, at any component (or stage) in the assessment process. The overall quality of the assessment system may therefore be limited by the quality of the weakest link. Newton and Shaw's (2014) framework for the evaluation of assessment and assessment systems is also useful because it provides a cogent structure through which to evaluate the quality of assessments and assessment systems that bypasses the terminology contests which cloud current discourse about evaluation of assessment quality. Newton and Shaw's framework distinguishes between the technical quality and social value of assessment, and provides a structure through which to evaluate both. Technical quality encompasses the assessment instrument and practice as mechanisms for assessing the target learning outcomes, the intended uses of primary decisions that will be made based on the assessment results, and the intended impact of these decisions. Social value encompasses the acceptability of the anticipated and unanticipated costs, benefits, impacts, and side-effects of the assessment or assessment system. The framework enables both the technical quality and social value to be evaluated before an overall judgement of the acceptability of the assessment or assessment system be reached.

The ongoing lack of consensus over the definition of 'validity' is a barrier to clear and effective communication (Newton & Shaw, 2016). In this thesis, Newton and Shaw's suggestion of retaining the concepts encompassed by the various definitions but desisting from using the term entirely (either with its technical meaning, or in a non-technical way), has been taken. Therefore, instead of using the terms 'valid', 'validity', or 'validate', in relation to assessment, the intended meaning will be explained using explicit vocabulary.[23] A unitary definition of validity has not been selected and used in this thesis because to do so would risk decreasing the clarity of communication: If a reader is not aware of the definition being used, or assumes that a different definition is applicable, they may take a different meaning from the text than that which is intended. Similarly, the terms 'valid', 'validity', and 'validate' will not be used in relation to assessment with no technical meanings attached, because, if a reader does not realise this and instead ascribes technical meaning to them, the clarity of communication would be compromised.

2.1.4 Assessment methods and conditions

The ways in which evidence of students' achievement is generated and collected in assessment situations can vary with the assessment methods and conditions that are used

---

[23] For example, whether the assessment instrument assessed only the intended outcomes. Note that the terms 'reliable' and 'reliability' will be used in relation to assessment in the present thesis.

(Harlen, 2007). The range of assessment methods available is extensive, (e.g., written examinations, collection of evidence generated naturally from every-day experiences). Furthermore, assessments in most formats can vary regarding how restricted student responses are, from multiple-choice questions, to open-response assessment activities such as extended-response essays (McMillan, 2007; Miller et al., 2009). Assessment conditions (i.e., the time allowed and resources available to students as they undertake an assessment) also vary substantially. The time allowed may be tightly restricted (e.g., a speeded test) or extended (e.g., a portfolio compiled from a full year's work). Likewise, the resources allowed for reference or use during assessment vary enormously, from none at all to 'open book' (McMillan, 2007; Miller et al., 2009).

Different assessment methods and conditions allow for the appropriate assessment of different cognitive and affective domains or learning outcomes (Broadfoot, 2007; Harlen, 2007; Klenowski & Wyatt-Smith, 2014; McMillan, 2007; Miller et al., 2009). For example, the recall of factual knowledge in a specified scientific domain may be appropriately assessed via a timed, closed-book examination, but the application of a scientific investigation process would not be. Instead, the latter would be more appropriately assessed by having students conduct an actual scientific investigation. Furthermore, sound assessment requires the level of cognitive and affective demand of the assessment to be appropriate for the target learning outcomes (McMillan, 2007; Miller et al., 2009). In vocational education that aims to prepare students for a workplace, sound assessment requires that the assessment tasks, methods, and conditions be appropriate for that industry (Misko et al., 2014). In that context, assessment conditions may include the physical environment in which assessment evidence is generated or collected: How authentic to the normal workplace the conditions are impacts on how sound the assessment is (Crisp & Novaković, 2009; Misko et al., 2014; Vaughan, Gardiner, & Eyre, 2012). Sound, trustworthy assessment is contingent on the evidence generation and collection method(s), and the conditions under which that generation and collection occurs being appropriate to the assessed domain (Harlen, 2007).

Variation in assessment methods, including how restricted the allowable responses are, impact on reliability. There is a greater likelihood that assessors' decisions will lack reliability for less restricted responses because more judgement will be required (Harlen, 2007; Miller et al., 2009). This possibility highlights the challenge involved in assessing certain learning domains: Assessments that place less restriction on responses tend to enable appropriate assessment of more complex learning domains (McMillan, 2007; Miller et al., 2009). For example, authentic performance-based assessments with few restrictions on students' responses tend to be appropriate for the assessment of applied or complex knowledge and

skills, but are often challenged by lower reliability (Miller et al., 2009). In contrast, tests that emphasise the assessment of factual knowledge using closed items tend to have a high degree of reliability, but are only appropriate for a limited range or depth of cognitive domains (Broadfoot, 2007). In the assessment of vocational skills, this challenge is further heightened. Assessments that are authentic to a workplace tend to be context-bound and focused on multi-dimensional competence, and as such, may lack generalisability. Furthermore, assessors might assess small groups of students doing varied tasks in varied settings. These factors make ascertaining assessment reliability challenging (Crisp & Novaković, 2009).

2.2 Internal assessment

Assessments vary in terms of who[24] makes judgements regarding students' evidence, and whether that 'agent of judgement' is within the organisation providing the education under assessment, or external to it (Harlen, 2007). This is the fundamental distinction between 'internal assessment' and 'external assessment': In internal assessment, assessors are internal to the organisation (e.g., teachers); in external assessment, assessors are external to the organisation, (e.g., external examiners or a marking panel from the awarding body; Crooks, 2011). Internal assessment is known by various terms, including 'coursework' in the U.K.'s General Certificate of Secondary Education (GCSE; Crisp, 2013), 'Overall Teacher Judgements' for New Zealand's National Standards (Ward & Thomas, 2016), 'school-based assessment' (Klenowski & Wyatt-Smith, 2014), 'classroom assessment' (Brookhart, 2003; Lissitz & Samuelsen, 2007), and 'teacher assessments' and 'teacher judgements' (Harlen, 2007).

Across different internal assessment systems, the responsibility, control, and freedom that individual organisations have over assessment methods, conditions, and assessment materials vary (Harlen, 2007; Miller et al., 2009). Assessment materials may be developed by external bodies (e.g., common assessment tasks), in which case the individual organisations have no control over the materials. The methods and conditions may be prescribed by an external body, for the organisation to adhere to when developing the assessment materials (Harlen, 2007). Alternatively, organisations may have freedom to select assessment materials (e.g., from a bank of tasks or from a commercial provider), or to

---

[24] Or what (i.e., a person, or, in the case of some computer-based assessment, a software programme)

develop their own—in which case the organisation has the challenge and responsibility of selecting appropriate methods and conditions to enable the sound, trustworthy, and reliable assessment of targeted learning outcomes (Allal, 2013; Miller et al., 2009).

2.2.1 Assessor judgement process

Human judgement, and the judgement practice of those involved, are central to internal assessment (Wyatt-Smith, Klenowski, & Gunn, 2010). The process of making assessor judgements is cognitive (Adie, Klenowski, & Wyatt-Smith, 2012), broadly involving considering the student evidence, comparing that evidence to some type of reference point, and reaching a decision (Crisp, 2012, 2013; Wyatt-Smith et al., 2010). Yet, how assessors make their judgements about evidence tends to be hidden (Crisp, 2013; Klenowski & Wyatt-Smith, 2014), particularly when assessors operate alone, as is often the case for internal assessment. Studies by Adie et al. (2012), Crisp (2012, 2013, 2017), Wyatt-Smith and Klenowski (2013), and Wyatt-Smith et al. (2010), have illuminated how assessors make their judgements about student evidence and, in the quality assurance process of moderation, how moderation judgements are made about student evidence. The findings will be discussed after the studies are introduced.

Crisp's (2012, 2013, 2017) studies were conducted in the U.K. secondary school sector in the context of assessment and moderation of GCSE coursework, and explored the cognitive processes used by teachers (n = 13) when making assessor judgements, and by external moderators (n = 9) when making moderation judgements about marked coursework. Verbal protocol data (in which participants were asked to 'think aloud' while undertaking a task) were collected from participants as they individually marked items of student evidence or moderated marked coursework, allowing inferences to be drawn regarding participants' cognitive processes. For her 2013 study, Crisp interviewed the same sample of teachers as she did in her 2012 study, and then validated and expanded the interview findings via a survey of secondary teachers (n = 378) involved with coursework marking.

The studies by Wyatt-Smith et al. (2010), Adie et al. (2012), and Wyatt-Smith and Klenowski (2013), were set in the Queensland middle school sector in the context of social moderation, and analysed data from a large-scale Australian Research Council Linkage project (survey: n = 189; focus-teacher interview, audio, and observation data of face-to-face and online moderation meetings: n = 89; textual artefacts from moderation meetings; Adie et al., 2012; Wyatt-Smith & Klenowski, 2013). A limitation of these studies is that, at the time of the

Linkage project, Queensland had recently introduced a standards-based curriculum and approach to assessment, as well as common assessment tasks. Thus, most teachers were learning a new curriculum, and were new to standards-based assessment and moderation (Adie et al., 2012; Wyatt-Smith & Klenowski, 2013; Wyatt-Smith et al., 2010). Therefore, caution is needed when drawing inferences from the studies: The ways in which teachers approach judgement-making may change with practice, as they become more competent in standards-based assessment. Notwithstanding these issues, the studies offer the following insights into how teachers make judgements under standards-based assessment and moderation.

Teachers and moderators were found to display a common sequence of behaviours when marking student evidence (Crisp, 2012), or moderating school submissions (Crisp, 2017), beginning with orienting and reading, and then alternating between reading and evaluating aspects of the student evidence. Judgement practice was found to be complex, and to vary between people, in the approaches used to evaluating the evidence, and to making overall grade judgements (Adie et al., 2012; Crisp, 2012, 2013; Wyatt-Smith & Klenowski, 2013). Assessors and moderators were found to compare student evidence to explicit assessment criteria, and to other referents, including annotated exemplars and tacit knowledge, (e.g., internalised mental representations of the standards, criteria, and expectations of each grade, curricular knowledge, and recollections or actual examples of other students' work; Adie et al., 2012; Crisp, 2012, 2013; Wyatt-Smith & Klenowski, 2013; Wyatt-Smith et al., 2010). Professional experiences (including previous assessment and moderation experience, guidance and training, and collegial interactions) appeared to influence judgement practices (Crisp, 2013). Further, while judgement-making is a cognitive act undertaken by individuals, it also appeared to be a socially-situated practice (Adie et al., 2012; Crisp, 2013) in that it is influenced by understandings of criteria that are shared between assessors, as well as by working with, and learning from, colleagues.

2.2.2 Why internal assessment is used

Internal assessment is widely used across all levels of the education sector, internationally and in New Zealand. Through internal assessment, it is practicable to collect student evidence using a greater range of assessment methods and conditions than is typically possible in external assessment (Klenowski & Wyatt-Smith, 2014). The variety of methods and conditions feasibly available to internal assessment allows for knowledge, skills, and abilities to be assessed that are not amenable to assessment via written examination (S.

Johnson, 2013). As such, a broader range of learning outcomes can be appropriately assessed through internal assessment—and to a greater depth—than through external examinations (Harlen, 2007). The flexibility that characterises internal assessment has been a strong driver for its inclusion in various senior secondary school qualification systems (Crisp, 2013; Harlen, 2007). Internal assessment is also widely used in vocational education and training: As Crisp observed, "it is generally accepted as the best way of assessing certain skills" (2013, p. 128).

The flexibility of internal assessment allows organisations the freedom to select the assessment methods and timing to suit their students and situation (Hipkins, Johnston, & Sheehan, 2016). In the New Zealand secondary context, the last two New Zealand Council for Educational Research (NZCER) surveys[25] found growing support among teachers and principals for the flexibility afforded by NCEA to better meet students' learning needs (Hipkins, 2013; Wylie & Bonne, 2016). Furthermore, the 2012 survey found that almost three quarters of teachers believed that a range of assessment methods could be appropriate for NCEA assessments (Hipkins, 2013), which is arguably a requisite recognition to making use of the qualifications' flexibility to suit local contexts and students.

2.2.3 Properties of good assessment apply to internal assessment

The properties of sound and trustworthy assessment (refer 2.1.3) are as important in internal assessment as they are in external assessment: Assessment instruments must assess only what was intended, and at the correct academic level; there must be accuracy, consistency, and fairness in the assessor judgements made by different teachers, and in the judgements made over time (S. Johnson, 2013; Newton & Shaw, 2014). Additionally, because a core feature of internal assessment is the flexibility it affords—there can be a variety of assessment activities conducted in a variety of contexts, all assessing the same learning outcomes—there is, therefore, a need for *comparability* of these activities, i.e., equivalence, not 'sameness' (Crisp & Novaković, 2009).

---

[25] A national survey of New Zealand secondary schools has been conducted every three years since 2003 by the NZCER. The surveys were administered to principals, teachers, trustees, and parents associated with state and state-integrated secondary schools, and canvassed their perceptions about NCEA and other matters. The surveys provide a large body of data about the perceptions of key stakeholders regarding the national standards-based assessment and qualification system since its inception. The sample for each survey was large: *n* (2012) = 3209, comprising 177 principals, 1266 teachers, 289 trustees, and 1477 parents; n (2015) = 3433, comprising 182 principals, 1777 teachers, 232 trustees, and 1242 parents (Hipkins, 2013; Wylie & Bonne, 2016).

2.2.3.1 Portfolio-approaches to internal assessment

When aiming to assess complex or applied knowledge and skills, there is often a tension between collecting student evidence that accurately and authentically reflects those targets, and reaching reliable assessor judgements (Biggs & Tang, 2007; Broadfoot, 2007; Miller et al., 2009). In internal assessment, Newton (2007) suggested that the collection of appropriate assessment evidence is often emphasised at the expense of reliability, due to the nature of the target learning outcomes, and the variety of assessment methods, activities, contexts, and conditions used to assess them. This tension is especially evident in portfolio-approaches, in which the freedom to select student evidence to include for assessment compromises the reliability of assessor judgements and poses challenges for achieving consistency in moderation judgements. The studies of Black et al. (2010, 2011), Hipkins (2012), Thomas, Johnston, and Ward (2014), and Ward and Thomas (2016), all illustrate this tension.

Black et al. (2010, 2011) reported on a small-scale longitudinal intervention set in the lower secondary sector in England (*n* [teachers] = 18, from three schools). The aim was to enhance summative assessments in ways that also enhance student learning. The three-year, multi-stage intervention involved critique of existing assessment instruments and practices, their redevelopment and refinement, and making assessor judgements, through ongoing cycles of implementation and quality assurance. Data from a wide range of sources were analysed, including observations of summative assessments, moderation and other relevant meetings, individual and focus-group interviews, moderation materials, and participants' reflective diaries. Through the intervention, the researchers observed the tension between flexibility and comparability that is inherent in portfolio-approaches to assessment. Measures that were subsequently implemented to enhance comparability in assessor judgements while maintaining the flexibility to select appropriate student evidence from different contexts, included specification of common evidence components for inclusion, stipulations regarding portfolio composition, and the implementation of robust intra-school and inter-school quality assurance (moderation) processes (Black et al., 2010, 2011). However, the effectiveness of those measures for increasing reliability (between assessors or over time) was not evaluated.

In their large-scale empirical study,[26] Thomas et al. (2014) examined the relationships between the main literacy and numeracy measures used in New Zealand's secondary and tertiary sectors, including the Learning Progressions for Adult Literacy and Numeracy[27] (learning progressions; measured via the Literacy and Numeracy for Adults Assessment Tool[28]), and the then newly-introduced internally assessed unit standards that explicitly address literacy and numeracy competencies. The unit standards were designed to align with the learning progressions (literacy: step 4; numeracy: step 5), and to be assessed against a portfolio of evidence generated naturally from students' normal learning programmes or everyday life. Thomas and her colleagues found that almost all of the Year 11 and 12 students in their study who had met the NCEA literacy or numeracy requirements via the unit standards were at least one step below the required level on the relevant learning progressions, and more than half were at least two steps below. The findings suggested that certification of students' literacy and numeracy levels by the unit standards did not reliably indicate those students' literacy or numeracy competencies. The problems found were systemic. However, it also seems likely that between-assessor inconsistencies in the evidence selected for assessment and in assessor judgements also contributed to the discrepancies: The unit standards were internally assessed and portfolio-based. Furthermore, at the time of data collection, the unit standards were new, and the alignment with the learning progressions was not explained in the unit standard-documents themselves (NZQA, 2010a, 2010b). Another possible contributor was suggested by Hipkins (2012): In her case study of one school's system for managing the assessment of the literacy unit standards, Hipkins observed that, although teachers of other subjects might collect student evidence for inclusion into students' literacy or numeracy assessment portfolios, not all teachers were likely to have the requisite literacy or numeracy expertise to make accurate judgements about the quality of that evidence. As such, the reliability of assessment judgements for those standards may have been compromised.

A final example pertaining to the challenges involved with portfolio-based assessments also comes from the New Zealand context. In a large-scale study into the implementation of National Standards from introduction (in 2010) through 2014, Ward and Thomas (2016) evaluated the reliability of assessor judgements (which were based on a portfolio-approach) against the National Standards. The researchers concluded that there was a widespread lack of consistency in assessor judgements against the National Standards, between school

---

[26] Their study used two sets of student data: 1) *n* (tertiary) = approx. 36,000 students aged 16-20 years; 2) school sample (from 15 schools): *n* (year 8) = approx. 1,000; *n* (year 11) = approx. 800; *n* (year 12) = approx. 800 (Thomas et al., 2014).

[27] Tertiary Education Commission (2008a, 2008b)

[28] Tertiary Education Commission (2016)

types, and over time. For each finding, Ward and Thomas considered alternative explanations, but concluded that inconsistency in judgements between teachers was the most likely explanation, and that this inconsistency was "unsurprising" because the National Standards were a new initiative, and resources were still being developed to support teachers to make consistent judgements (2016, p. 2). However, it seems likely that other aspects also contributed to the lack of consistency found, including the portfolio-approach to assessment underpinning those judgements, and the freedom afforded to teachers regarding evidence selection for those portfolios. In addition, it appears that robust intra-school and inter-school moderation in which all schools engaged was lacking (Ward & Thomas, 2016; refer also 2.4.3.5).

Because the results of high-stakes summative internal assessments are used to make important decisions (e.g., whether students are awarded a qualification), it is critical that the assessments are of high quality. It is necessary for organisations to implement processes to ensure that internal assessments conducted under their own auspices are sound and trustworthy. Regulatory authorities and quality assurance bodies (e.g., NZQA, and England's Office for Standards in Education, Children's Services and Skills; Altrichter & Kemethofer, 2015) also have an interest in the quality of education that students receive, and of internal assessments (and the credibility and legitimacy of assessment results), and so implement quality assurance frameworks to address these matters (Broadfoot, 2007; Ehren, Altrichter, McNamara, & O'Hara, 2013; Harlen, 2007; Misko, 2015b).

2.3 Quality assurance in education

In the context of the education sector, quality assurance is the systematic review and evaluation of educational services to ensure that the expectations and requirements of a governance regime are met and maintained (Altrichter & Kemethofer, 2015). The overarching aims of quality assurance systems are to monitor, assure, and improve the quality of education and qualifications (Ehren et al., 2013), and to create and maintain trust in the performance of organisations and the education system (European Association for Quality Assurance in Higher Education et al., 2015). Scholars including Broadfoot (2007), Dill (2007), and Harvey and Newton (2007) have argued that governance regimes also use quality assurance as a policy instrument to steer the education sector.

2.3.0.1 The concept of quality

Quality, as a concept, is subjective and relative. It means different things to different people, and sometimes, different things to the same people, depending on the circumstances or situation (Harvey & Green, 1993; Kleijnen, Dolmans, Willems, & van Hout, 2013; K. Maguire & Gibbs, 2013; Van Kemenade, Pupius, & Hardjono, 2008; Wong & Li, 2010). It can also relate to either processes, or outcomes, or both (Harvey & Green, 1993). Different quality assurance approaches tend to engender different concepts of quality (Harvey & Newton, 2007; Harvey & Stensaker, 2008; K. Maguire & Gibbs, 2013; Van Kemenade et al., 2008). Furthermore, the concept(s) of quality to which stakeholders subscribe may influence their perceptions of a quality assurance regime (Harvey & Stensaker, 2008; Kleijnen et al., 2013; Kleijnen, Dolmans, Willems, & van Hout, 2014). In quality assurance regimes in which there are strong accountability drivers, the concepts of quality that are commonly evident reflect Harvey and Green's (1993) concepts of quality as 'consistency', 'meeting set standards', 'fitness for purpose', and 'value for money' (Broadfoot, 2007; Harvey & Newton, 2007; Harvey & Stensaker, 2008). Quality can be seen as consistently meeting certain standards or criteria, whether the standards or criteria have been set within an organisation, or externally (e.g., by a regulatory authority). This concept of quality allows for different standards in different organisation types and enables transparency because the required standards can be communicated to stakeholders. Being fit-for-purpose also constitutes quality, bearing in mind that different stakeholders (e.g., teachers, students, funders, employers) may value different purposes. Value-for-money is a concept of quality that is linked to efficiency and effectiveness in terms of use of resources. The notions of quality as consistently meeting set criteria, being fit-for-purpose, and being value-for-money, appear to underpin NZQA's external quality assurance regime that is the subject of the present study (e.g., NZQA, 2014c, 2014d). However, it is acknowledged that participants in the present study were situated in different organisations, each with its own internal quality assurance system, and in these different contexts, other conceptions of quality may have been evident.

2.3.0.2 Common elements of quality assurance frameworks

There are common elements in most quality assurance frameworks, although the processes and approaches vary in different jurisdictions and parts of the education sector. Commonly, expectations for performance or process (e.g., quality standards, criteria or performance

indicators) are established[29] and communicated to organisations and stakeholders. Organisations are evaluated against those expectations via various quality assurance processes (e.g., accreditation). The evaluation findings are communicated to stakeholders and the organisations themselves and are then used by the various parties to inform subsequent actions. The consequences of the evaluation (e.g., rewards, sanctions, or interventions imposed in response to the evaluation findings) provide accountability (Altrichter & Kemethofer, 2015; Ehren et al., 2013). The quality assurance frameworks themselves are situated within the political and cultural contexts of the countries, organisations, and professions in which they occur, and as such, the aspects covered by the standards and indicators are generally tailored to those contexts (Misko, 2015a).

The requirements and components of quality assurance frameworks are set by regulatory authorities or quality assurance bodies at national, provincial, district, or municipal levels (Ehren et al., 2013). Internationally, it is commonplace for quality assurance frameworks to contain both internal and external components as mandated requirements (Academic Quality Agency for New Zealand Universities, 2013; Croxford, Grek, & Shaik, 2009; Education Review Office, 2014; Ehren et al., 2013; Lillis, 2012; Misko, 2015a, 2015b; Ng, 2007; Vanhoof, Van Petegem, Verhoeven, & Buvens, 2009; Wong & Li, 2010). It is widely assumed that both internal and external aspects are necessary to fulfil the accountability and improvement functions of a quality assurance system (e.g., by Danø & Stensaker, 2007; Law, 2010; Meade & Woodhouse, 2000; Misko, 2015a). In the context of the present study, the external quality assurance frameworks in place in the secondary and tertiary sectors include the requirements for internal quality assurance by each organisation and engagement in external quality assurance (Education Review Office, 2014; Misko, 2015b; NZQA, 2014f). However, as Braun et al. (2011) have observed, policies and regulations are rarely implemented as envisioned by policy makers. Instead, individual organisations respond to and enact policy and regulatory requirements (including those pertaining to external quality assurance regimes) in contextualised, situated, creative, and diverse ways. The process (and resulting enactment) involves interpretation and translation, and is influenced by a wide range of organisational factors, including the context and culture, existing conditions, commitments and practices, staffing levels and expertise, and external pressures and expectations (Braun et al., 2011). It seems likely that the New Zealand environment of self-governing and self-managing education organisations (Ministry of

---

[29] Those quality standards or criteria that are set by regulatory authorities or quality assurance bodies can be considered to be 'regulatory' standards, because registered organisations must adhere to them; where those set by organisations themselves, can be considered to be 'voluntary' standards, because each organisation has control of them (Misko, 2015b).

Education, 2015; Openshaw, 2014) has produced greater levels of diversity in the enactment of policy than would occur in a more centralised governance and management environment.

2.3.0.2.1 Quality assurance components external to an organisation

External quality assurance components involve the evaluation of an organisation's performance or processes by parties external to the organisation, (i.e., external review). The external parties could be from a quality assurance or regulatory body, (e.g., auditors or evaluators; Misko, 2015a), or peers (e.g., colleagues from another school; Gilbert, 2012). External quality assurance processes, when implemented by the quality assurance or regulatory body, are used to enact and ensure organisations' adherence to the regulatory standards (Misko, 2015b). External review provides a level of independence from the evaluation judgements made within an organisation (i.e., self-evaluation), which is seen to be important to ensure an accountability function (Misko, 2015a). External review can add transparency to the quality assurance system, which is also important for accountability (Cuttance, 2005; Danø & Stensaker, 2007; European Association for Quality Assurance in Higher Education et al., 2015). Peer review is assumed to assist in enhancing quality by supporting organisations in their improvement efforts and providing professional learning opportunities for those involved (Ehren, Perryman, & Spours, 2014; Gilbert, 2012; Misko, 2015a; Stoten, 2012). Through the course of conducting this literature review, little evidence was found that evaluated the effectiveness or impact of either external review or peer review in quality assurance regimes, apart from the research in which external review and peer review occurs in a moderation context (refer 2.4.3, 2.4.4).

2.3.0.2.2 Quality assurance components internal to an organisation

Internal quality assurance occurs within an organisation. In the literature, internal quality assurance is variously called 'self-assessment' (e.g., by Cardoso, Rosa, & Santos, 2013; Montecinos, Madrid, Fernández, & Ahumada, 2014), 'self-evaluation' (e.g., by Gustafsson et al., 2015; Vanhoof et al., 2009; Wong & Li, 2010), and 'self-review' (e.g., by Education Review Office, 2014; and Gilbert, 2012). Henceforth, self-assessment, self-evaluation, and self-review, are referred to as 'self-evaluation'. When quality assurance processes are conducted within an organisation, the focus of the evaluation can be at organisation, department, or individual levels. Self-evaluation is thought to play a central role in educational improvement by stimulating self-reflection, learning, and a focus on

improvement (Danø & Stensaker, 2007; Misko, 2015a). However, much of the evidence supporting the centrality of self-evaluation's role is perception-based: including assumptions that underpin policy (e.g., Croxford et al., 2009; Ehren et al., 2013; Ehren, Perryman, & Shackleton, 2015; Montecinos et al., 2014), and self-report data of stakeholder perceptions (e.g., Gustafsson et al., 2015; and Vanhoof et al., 2009). A rare example of research that used another data source alongside stakeholder perception data was Wong and Li's (2010) mixed methods study into the Hong Kong early childhood education sector, which drew on published external quality assurance inspection reports. While their study suggested that self-evaluation was not well established in that sector, the researchers found that there was a positive correlation between kindergartens' performance in self-evaluation and in Learning and Teaching: the better the kindergartens performed in self-evaluation, the fewer problems were found with their performance in Learning and Teaching performance indicators. Although it was inappropriate to conduct parametric tests (given the ordinal data), an appropriate non-parametric test was used to ascertain the correlation between the two variables, meaning that Wong and Li's results showing a positive correlation between the variables were valid.

2.3.1 Quality assurance processes that indirectly address internal assessment

Various quality assurance processes that are implemented in the education sector address the quality of internal assessment in passing, although it is not their primary focus. These include initial entry processes (e.g., accreditation and approval), recurrent processes (e.g., inspection or audit), and outcome-focused processes (e.g., evaluative reviews). Misko (2015b) observed coherence between the criteria applied in the various quality assurance processes in many of the quality assurance regimes she examined, which allowed the different processes to reinforce each other. The examples of quality assurance processes described below are those that are the most relevant to the context of the present study.

2.3.1.1 Internal entry processes

Initial entry quality assurance processes, including accreditation and approval (and, in the New Zealand context, PTE registration, and consent-to-assess) are implemented before an organisation can deliver education or training, assess against certain assessment standards,

or award qualifications (Harlen, 2007; Misko, 2015a, 2015b; NZQA, 2013a, 2016a). In most jurisdictions, these front-end processes are conducted by regulatory authorities or delegated bodies (Misko, 2015a), although programme approval may also be conducted internally. Through each process, the organisation, programme, or course is evaluated to determine whether the relevant criteria have been met. The criteria cover things such as ensuring that the organisation has the capability and resources to sustainably deliver, assess, and quality-assure programmes of learning or assessment standards, and that programmes are educationally coherent, appropriately focused, and will be assessed appropriately (Misko, 2015a; NZQA, 2011c, 2013a, 2016a).

2.3.1.2 Recurrent process-focused processes

Process-focused recurrent quality assurance processes, including inspection and audit, are concerned with 'inputs' (e.g., whether teaching staff have appropriate training), and practices and systems (e.g., pedagogical approaches, assessment practices; NZQA, 2009). They are implemented periodically and are used to check an organisation's ongoing adherence to, and implementation of, quality standards or criteria (Gustafsson et al., 2015; Misko, 2015a). They can be conducted within an organisation as part of an internal quality assurance regime, or by parties external to an organisation as part of an external quality assurance regime. (Thus, the results of the inspection or audit may or may not be communicated to external stakeholders.) They may involve an analysis of an organisation's documentation and data, or site visits (Ehren et al., 2013; Ehren & Hatch, 2013; Harlen, 2007; Misko, 2015b). Inspections and audits vary in scope and focus, and the quality standards or criteria applied vary from being broad and wide-ranging, to narrow and targeted (Ehren et al., 2013; Gustafsson et al., 2015; Misko, 2015b). As such, not all audits or inspections would consider internal assessment matters. However, NZQA's Managing National Assessment (MNA) audits of New Zealand's secondary schools do always consider internal assessment, as their focus is on how each school manages summative assessment for national qualifications (Hipkins et al., 2016). The results of the MNA audits are published on the NZQA website (NZQA, 2014a). Historically, external audits were a central component in the external quality assurance regime in New Zealand's non-university tertiary sector (NZQA, 1992a). Until an evaluative approach to quality assurance was adopted in 2009 (NZQA, 2012a), NZQA conducted the external 'quality audits' of PTEs (NZQA, 1992b). ITPs were externally audited by the Institutes of Technology and Polytechnics Quality, as the then quality assurance body with delegated authority from NZQA (Institutes of Technology and Polytechnics Quality, 2006). Under the current evaluative framework external audits are not standard practice,

although NZQA does use inspection in investigations into suspected non-compliance with legislation, or NZQA rules, regulations, or consents (NZQA, 2013a).

2.3.1.3 Outcome-focused processes

Outcome-focused quality assurance processes, including evaluative review, consider the quality of educational outcomes. Through evaluative review, an organisation's performance is periodically assessed against set evaluative questions (NZQA, 2012a). Key inputs and processes (e.g., the quality of teaching or assessment practices) are considered only in as much as they contribute to, or have an impact on, outcomes (NZQA, n.d.e). Evaluative reviews may be conducted within an organisation itself or by parties external to the organisation (e.g., reviewers from the relevant quality assurance body), and thus, the results may or may not be made public. The scope of each evaluative review is determined by the evaluative questions and the focus areas (e.g., programmes of learning) of the review (Education Review Office, 2014; NZQA, n.d.e). Thus, internal assessment and moderation are not always considered, and when they are, it tends to be only in terms of their contributions to, or impact on, the evaluative questions.

2.3.1.3.1 Outcome-focused processes used in the NZ context

In the New Zealand context, evaluative review is a central plank in the external quality assurance regimes that apply to the compulsory education and non-university tertiary education sectors, and the results of the external reviews are published (Education Review Office, 2014; NZQA, 2012a). The evaluative quality assurance framework that applies to the tertiary sector requires organisations to conduct their own evaluative reviews (in that context, called 'self-assessment'), as well as to be reviewed by evaluators from NZQA (called external evaluative review—EER; NZQA, 2014f). New Zealand's Education Review Office utilises a similar evaluative quality assurance framework with state schools, which comprises both the school's own internal evaluative reviews (called 'self-review') and external evaluative reviews (Education Review Office, 2014). Evaluative review has been promoted by the New Zealand quality assurance bodies that use it as evidence of being responsive to the unique contexts of individual organisations (Education Review Office, 2014; NZQA, 2012a), and of allowing reviewers to recognise that what quality 'looks like' can be context-dependent (NZQA, n.d.e). However, a tension exists between the flexibility afforded by the approach (whereby what is judged to be quality may 'look different' in different contexts) and

achieving consistency in evaluative review judgements and feedback, reflecting the challenges in ensuring reliable portfolio assessments (refer 2.2.3.1). An independent review into NZQA's evaluative quality assurance framework found dissatisfaction within the tertiary sector about a perceived lack of consistency in published review reports (feedback and awarded ratings) and of the capability of evaluators. The independent panel signalled this finding as a threat to the credibility of, and confidence in, the system (NZQA, 2012a).

Since the data for the present study were collected, another outcome-focused quality assurance process was introduced into New Zealand's tertiary education sector: consistency reviews. These periodically evaluate the consistency of graduate outcomes of the various programmes leading to a single New Zealand qualification, to ensure a measure of comparability between graduates of the programmes in terms of meeting the qualification outcomes (NZQA, 2014b). Consistency reviews are outside of the scope of this thesis and are not discussed further.

2.4 Moderation of internal assessment

Moderation is a process that directly addresses internal assessment (Harlen, 2007). It is used in many jurisdictions for quality assurance, including Australia and Canada (Klenowski & Wyatt-Smith, 2014), England, Wales and the Netherlands (S. Johnson, 2013), Scotland (Grant, 2012), New Zealand (Hipkins et al., 2016), and Switzerland (Mottier Lopez & Pasquini, 2017). Moderation is used across most levels of education, including primary (Wyatt-Smith et al., 2010), secondary (Crisp, 2017; Hipkins, 2013), and tertiary, including vocational education and training (Halliday-Wynes & Misko, 2013), workplace training (Vaughan et al., 2012), and university (Adie et al., 2013). In different contexts, moderation differs in the component (or stage) of assessment addressed, focus, approach, implementation, and function. These variations are now described. Greatest consideration is given to the two main groups of functions of moderation: accountability, quality control, and maintaining public confidence, and organisational and professional improvement and learning (Adie, 2012; Bloxham, Hughes, & Adie, 2016; Crisp, 2017; Ehren et al., 2013; Grant, 2012; Harlen, 2007; Klenowski & Wyatt-Smith, 2014; Misko, 2015a; Wong & Li, 2010).

2.4.0.1 Different components of assessment process addressed by moderation

Moderation can be used to address various components of the assessment process as described by the different links in Crooks and his colleagues' (1996) assessment chain: from the conceptualisation and design of assessment instruments to the grades awarded to students. The focus of moderation differs across assessment components (Bloxham et al., 2016). The two components of the assessment process that moderation most commonly addresses are assessment instruments and assessor judgements.

When moderation addresses an assessment instrument, its focus is on the extent to which the task can be used to collect student evidence of the targeted performance or learning, and whether the instrument assesses *only* what is intended (as defined by the target learning outcomes or standards), and with no extraneous requirements (Bloxham et al., 2016). Moderation that focuses on assessment instruments is sometimes called 'validation' (Misko et al., 2014), 'pre-moderation'[30] (Afrin, 2011), or 'pre-assessment moderation'[31] (NZQA, 1996). Misko et al. (2014) found that in the Australian vocational education and training sector, stakeholders from relevant industries were sometimes involved in moderation of assessment instruments, to ensure that instruments, methods, and conditions were fit-for-purpose and appropriate from an industry perspective. Misko et al. also reported that participants in their study perceived there to be a stronger focus in their sector on moderation of assessment instruments than of assessor judgements.

When moderation addresses assessor judgements and grades awarded, its focus is the accuracy, consistency, and fairness of those judgements (Klenowski & Wyatt-Smith, 2014). For example, the external moderation of GCSE coursework in England appears to only address assessor judgements and provisionally-awarded grades (Crisp, 2017). Likewise, in the context of New Zealand's Year 1-8 National Standards, Ward and Thomas (2016) solely addressed assessor judgements when referring to the aims of moderation in their research study.

Moderation can also address other components of the assessment process. For example, in the intervention of Black et al. (2010, 2011; refer 2.2.3.1), moderation also addressed the

---

[30] Anecdotally, the term 'pre-moderation' is also sometimes used in reference to exercises to calibrate assessors' judgement prior to marking student evidence.
[31] Anecdotally, the term 'pre-assessment moderation' is used within some organisations even when the moderation of assessment instruments occurs after the assessment has been conducted.

conceptualisation of assessment and the selection of assessment methods. It involved participants collaboratively conceptualising and designing assessment instruments that enabled summative assessment of the target learning outcomes and supported student learning. Bird and Yucel (2013; refer 2.4.3.4) reported on an intervention in an Australian university in which moderation occurred sufficiently early in the course delivery to inform teaching, to enable tutors to better prepare students for assessment, and to enable students to make better use of assessment feedback received to improve their own learning. Moderation can also address assessment implementation. For example, Vaughan et al. (2012; refer 2.4.4) reported in a workplace training context that moderation was conducted as a 'ride-along', whereby a moderator accompanied an assessor when visiting apprentices, and observed assessments occurring. Moderation may also be conducted prior to the marking of assessments, to calibrate assessors' judgements and ensure that they have an accurate understanding of the assessment criteria and standards before marking student work. For example, the moderation initiative that Crimmins et al. (2016) reported on in an Australian university included a pre-marking moderation component that focused on calibration.[32] Scholars including Beutel, Adie, and Lloyd (2017), and G. Johnson (2015), have argued cogently that in order to realise its potential to enhance student learning and outcomes, moderation should address different stages throughout the teaching, student learning, and assessment cycle, instead of being a 'one-point-in-time' practice (as it is when solely addressing assessor judgements already made).

2.4.1 Different moderation approaches

Different approaches to moderation are used in different contexts. The approaches generally fall into three main categories—expert, social, and statistical—although hybrid models are used in many situations.

Expert moderation involves an 'expert' in the role of moderator, who evaluates the assessment instrument, samples of student work and associated assessor judgements, or assessment implementation (Crisp, 2017; Klenowski & Wyatt-Smith, 2014). The moderator makes a judgement regarding the quality of the instrument, judgements, or implementation,

---

[32] The multiphase initiative involved casual academic staff (n = unreported, but drawn from the 18 casual tutors) in a first-year communications course, and was used to meet accountability requirements and provide professional development. Along with the calibration workshop, two phases of expert moderation were involved, including post-marking feedback to individual tutors. Data were collected via observation of the calibration workshop, and focus-group interviews. Refer also 2.4.1, 2.4.4.2, and 2.4.5.1.

and then provides feedback (usually to the assessing teacher or organisation). For example, Crisp (2017) reported that expert moderation was the primary approach used by examination boards in England for GCSE coursework. The moderation initiative in an Australian university that Crimmins et al. (2016; 2.4.0.1) reported on included two phases of expert moderation. In the first, a sample of marked work from each tutor was submitted for moderation, and feedback was provided to that tutor about their judgements. The other phase occurred prior to the final approval of grades, in which a random selection of assessed work was moderated. Variations on the expert moderation approach include panel moderation (involving a panel of experts; Queensland Studies Authority, 2010), and peer moderation (involving a colleague in the role of moderator; Bloxham et al., 2016).

Social moderation (sometimes called 'consensus moderation') involves a group of teachers meeting to consider, discuss, and reach agreement about assessment instruments, standards, or assessor judgements for samples of student work (Hipkins & Robertson, 2011). There are variations in the implementation of social moderation, but common elements comprise discussion, debate, and negotiation between participants, with a focus on reaching agreement and developing shared understanding of the matters at hand (e.g., the requirements and application of standards, or the qualities of evidence that meet the assessment criteria or grade levels; Grainger et al., 2016; Sadler, 2012). Variations in the implementation of social moderation include the structure and focus of moderation meetings, whether meetings are facilitated, and whether common activities (e.g., grading samples of student work) are completed individually by participants prior to meeting for discussion at the meeting (Klenowski & Wyatt-Smith, 2014).

Statistical moderation focuses only on the grades awarded. Statistical procedures are used to compare internal assessment grades to another measure (e.g., external examination results), and the former are then typically adjusted to bring them into alignment with the latter (Bloxham et al., 2016; Crisp, 2017; G. Johnson, 2015). For example, the external moderation system for the Western Australian Certificate of Education (WACE) includes a statistical moderation component (Curriculum Council, 2007), and in Queensland, tertiary entrance ranks are determined using the external Queensland Core Skills Test results to statistically moderate the results of the externally-moderated, internally-assessed, Queensland Studies Authority senior secondary subjects (Klenowski & Wyatt-Smith, 2014).

Variations in the implementation of moderation also include the sampling approaches used for selection of assessed work to be submitted for moderation. Examples of sampling approaches include random selection (NZQA, 2013b), selecting samples that are close to or

on certain grade boundaries (Ward & Thomas, 2016), selecting a range of samples to show distributions of grades (Beutel et al., 2017), selecting samples that are considered to be 'sound' examples of certain grades (Crimmins et al., 2016), and selecting samples that are considered outliers or anomalous in terms of the distribution of grades provisionally awarded (Bloxham et al., 2016). Further differences in implementation include whether moderation can result in changes to the formal grades awarded to students. For example, external moderation for the GCSE in England is used to confirm and approve the grades awarded, or where necessary, to alter them or require the student work to be remarked (Crisp, 2017). In contrast, national moderation as conducted by NZQA in New Zealand generally has no direct impact on the actual grade awarded to students (NZQA, 2014a, 2014c).

2.4.2 Moderation within organisations, between organisations, and at jurisdiction level

Internal moderation is conducted within organisations to quality-assure internal assessment and is usually specified as a part of an organisation's quality management system. In many jurisdictions, organisations are required by regulation to have a system of internal moderation in place if they are to conduct summative assessment (Bloxham et al., 2016; Misko, 2015a). In the New Zealand context, this requirement is specified in multiple places, including through accreditation requirements (NZQA, 2011c), consent-to-assess requirements (NZQA, 2011b), and assessment rules for schools (NZQA, 2014a), although the requirement to implement internal moderation does not extend to specifying how internal moderation is to be conducted. In the Australian higher education context, the introduction of a regulatory requirement compelling universities to report their moderation practices appears to have prompted a slew of research into internal moderation practices in that sector— including Adie et al. (2013), Beutel et al. (2017), Bloxham, Hudson, den Outer, and Price (2015), Bloxham et al. (2016), Crimmins et al. (2016), and Grainger et al. (2016). Internal moderation systems generally involve social moderation, or expert (or peer) moderation, or a combination of approaches (Crisp, 2017; G. Johnson, 2015; Klenowski & Wyatt-Smith, 2014). For example, the university initiative reported by Crimmins et al. (2016) involved a hybrid model: Social moderation was used to calibrate assessors' judgement prior to conducting any marking, and two phases of expert moderation were conducted (during and following marking; refer 2.4.1).

Moderation may also be conducted between organisations (Grant, 2012; Klenowski & Wyatt-Smith, 2014). As with internal moderation, this can use an expert or peer moderation approach, or a social moderation approach (Hipkins & Robertson, 2011). In situations in which there is only one teacher of a learning domain in an organisation, the requirement for internal moderation may be fulfilled by conducting moderation with a teacher from another organisation (NZQA, 2011b). Conducting moderation among teachers from clusters of organisations is also used to assist with ensuring comparability between organisations in terms of assessment approach and assessor judgements (Grant, 2012; Klenowski & Wyatt-Smith, 2014; Mottier Lopez & Pasquini, 2017). Ensuring comparability between organisations may also be required by regulatory authorities. For example, a multi-institutional social moderation initiative involving nationally-set accounting threshold standards, as reported by Watty et al. (2014; refer 2.4.4.1) and O'Connell et al. (2016; see 2.4.3.4), occurred in the context of changes to the regulations governing the Australian university sector that included the requirement for national comparability against published disciplinary standards.

Moderation is also conducted by relevant authorities at a state or national level (Crisp, 2017). This is known as 'external', 'jurisdiction-wide', or 'national' moderation, and is conducted against jurisdiction-wide assessment standards or syllabi (Crooks, 2011; Misko, 2015b). It functions to ensure comparability of assessment and maintenance of standards at a jurisdiction-level (Crisp, 2017), and as such, tends to have an accountability purpose (see 2.4.3). This type of moderation is generally part of the external quality assurance regime in place within that jurisdiction. In the New Zealand context, the national moderation as conducted by NZQA (henceforth called 'NZQA moderation') for the NCEAs is conducted using an expert moderation approach (Hipkins et al., 2016), while, in Australia, the external moderation for the WACE utilises expert, social, and statistical, moderation approaches (Curriculum Council, 2007). For organisations, engagement in the relevant external moderation is generally a requirement of assessing the associated standards (e.g., NZQA, 2011b).

2.4.3 Moderation for accountability purposes

In the eyes of the public, the legitimacy of an assessment system generally hinges on its reliability, and this tends to be a key concern for governments (Broadfoot, 2007). When used for high-stakes purposes, internal assessments often suffer from lower levels of public confidence than external assessments (Harlen, 2007). Common perceptions of internal

assessment that lead to decreased public confidence include the use of inappropriate assessment tasks, leniency in marking, and plagiarism (S. Johnson, 2013).

In many jurisdictions, authorities require organisations to engage with and implement moderation and other quality assurance processes to ensure that the quality of students' achievement and assessors' judgements are acceptable, that graduates have the knowledge and competencies certified by their qualifications, and that stakeholders can have confidence in this (Controller and Auditor-General, 2012; European Association for Quality Assurance in Higher Education et al., 2015; Halliday-Wynes & Misko, 2013). Specifically, moderation is used to ensure the quality and integrity of internal assessments and the use of robust, fair, and credible assessment processes, through which to uphold the credibility of the associated assessment standards and qualifications (Adie et al., 2013; Crisp, 2017; Crooks, 2011). Thus, accountability, and the associated functions of quality control, and maintaining public confidence, are key purposes for which moderation and other quality assurance processes are used (Ehren et al., 2013; European Association for Quality Assurance in Higher Education et al., 2015; Halliday-Wynes & Misko, 2013; Harlen, 2007; Klenowski & Wyatt-Smith, 2014; Misko, 2015a; Wong & Li, 2010). Quality assurance systems across all levels of education, from early childhood education (e.g., Wong & Li, 2010) to university (e.g., Adie et al., 2013), have these aims. Scholars including Crisp (2017), Ehren and Hatch (2013), Klenowski and Wyatt-Smith (2014), and Sadler (2012) have asserted that accountability is the main function and most common goal of moderation and other quality assurance processes (although not the only function or goal, as returned to in 2.4.4). Adie et al. (2013) referred to the discourse of accountability as the 'predominant discourse' regarding moderation.

2.4.3.1 Accountability, quality control, and maintenance of public confidence

Fundamentally, accountability is a relationship between two parties in which one party is responsible for their actions and performance, and is obliged to account for these to the other party (Gilbert, 2012; Harlen, 2007). In the context of internal assessment, the actions and performance that organisations (e.g., schools) or individuals (e.g., teachers) are accountable for could include the quality of the teaching, student learning, and assessment experiences provided, levels of student achievement, integrity of assessment instruments or practices used, use made of results, or compliance with policy and regulatory requirements

(Gilbert, 2012; Harlen, 2007; Harvey & Newton, 2007; Klenowski & Wyatt-Smith, 2014). The stakeholders to which the first party might be accountable include students, parents and caregivers, the local or wider community, the organisation's governing body, funders, employers, other education organisations, or the government, quality assurance bodies, or regulatory authorities (Gilbert, 2012; Harlen, 2007; Harvey & Newton, 2007). Common components of an accountability relationship reflect the common elements of quality assurance frameworks, including specified expectations or minimum requirements for the first party's actions and performance (as set by an external authority or by the organisation itself), associated incentives, sanctions, or interventions (that come into effect depending on the first party's performance in relation to those expectations), and reporting of the first party's performance to stakeholders in relation to expectations (Altrichter & Kemethofer, 2015; Broadfoot, 2007; Ehren et al., 2013; Gilbert, 2012; Gustafsson et al., 2015; Harlen, 2007; Klenowski & Wyatt-Smith, 2014).

Quality control is closely associated with, and contributory to, accountability. Quality control concerns evaluating the actions or performance that organisations or individuals were accountable for, and ascertaining whether the expectations or minimum requirements were met (Ehren et al., 2013). Quality control focuses on the maintenance of quality standards and adherence to procedural requirements (Van Kemenade et al., 2008). Thus, quality control—or the perception of it—enables accountability.

The maintenance of public trust and confidence in the education system is thought to result from accountability. The public or stakeholders are the second party in an accountability relationship; that is, they are stakeholders to whom an organisation or individual is accountable. Through quality control and the resulting accountability (and the transparency and demonstration of quality that these provide), the widely-held educational policy assumption[33] is that public confidence and trust in an education system, the education provided, and the qualifications awarded, can be maintained (Broadfoot, 2007; European Association for Quality Assurance in Higher Education et al., 2015; Harlen, 2007; Klenowski & Wyatt-Smith, 2014).

---

[33] Although O'Neill argued that this assumption is sometimes misplaced: When accountability systems incentivise perverse behaviours and outcomes that damage student learning (i.e., a core purpose of education), those systems can be "a source rather than a remedy for mistrust" (O'Neill, 2013, p. 10).

2.4.3.2 Key drivers in quality assurance regimes

The demand for accountability, quality control, and the maintenance of public confidence, have been key drivers in the enactment of quality assurance regimes in most jurisdictions (Altrichter & Kemethofer, 2015; Ehren et al., 2013; Ehren & Hatch, 2013; Gilbert, 2012; Harvey & Newton, 2007; Misko, 2015b; Montecinos et al., 2014). In terms of national moderation, this is the case for the secondary school qualifications mentioned earlier (the GCSE, NCEAs, and WACE). It is also the case in internal moderation across the levels of education, from primary (Ward & Thomas, 2016), to university (Beutel et al., 2017). Furthermore, the political importance of maintaining public trust and confidence in education and qualifications systems has been demonstrated[34] by New Zealand's recent abandonment of the Years 1–8 National Standards (Gerritsen, 2017d; Moir, 2017). Ongoing changes to the coursework component of the U.K.'s GCSE provide one international example (for accounts, refer to S. Johnson, 2013, and BBC, 2013). The upcoming changes to the Queensland Certificate of Education (Queensland Curriculum and Assessment Authority, 2018) appear to provide another international example of changes made to an assessment and qualifications system, and the associated quality assurance regime, designed to ensure public confidence in the comparability of internal assessments and the qualifications those assessment results contribute towards. The political importance was also suggested by Misko's (2015a) finding that key functions of the quality assurance regimes in the vocational education and training sectors of 13 countries were to maintain public trust and confidence in the qualifications as awarded by each sector, and to answer—if not prevent—questions about the integrity of those qualifications.

2.4.3.2.1 Different requirements between NZ sectors suggest variation in policy drivers

In the New Zealand context, NZQA appears to have had, at least until recently, different levels of concern for maintaining public confidence in the assessment practices and qualifications awarded by secondary and tertiary sector organisations. One of NZQA's self-described roles is "to ensure that New Zealand's qualifications are regarded as credible and

---

[34] While it was arguably an ideological decision, the incoming 2017 Labour government cited a lack of confidence in the National Standards on the part of schools and parents as one reason for abandoning the mandatory use of the standards, and the NZEI (a teachers' union) and the New Zealand Principals' Federation both expressed their lack of confidence in the standards as accurate or reliable measures of students' performance (Moir, 2017). Wylie and Bonne (2014; discussed below) had also found very low levels of confidence in the reliability of the standards on the part of principals, teachers, and Boards of Trustees.

robust" (NZQA, n.d.d), for which maintaining public confidence in the qualifications system is critical. Issues of public confidence in the NCEA system—including its internal assessment components—have been key drivers of changes made to the system, including its moderation system, since its inception (Hipkins et al., 2016). As already noted (2.4.1), quality assurance of qualifications has not, as a matter of course, extended to approving or altering internal assessment grades awarded to students in either sector (NZQA, 2014a, 2014c). However, NZQA's concern for quality control, accountability, and maintaining public confidence, appears to have been stronger for the secondary than the tertiary sector, as suggested by differences in NZQA's requirements of organisations in each sector, and in NZQA's moderation practices.

NZQA requires that schools verify internal assessment results as accurate through internal moderation before they are reported to NZQA and entered into a student's Record of Achievement (NZQA, 2014a). In contrast, while tertiary education organisations are required to have moderation policies and procedures in place (NZQA, 2015, 2017b), to date there has been no stipulation that the accuracy of internal assessment results be verified before results are reported.[35] NZQA moderates two different sets of NCEA moderation submissions from the secondary sector annually (Hipkins et al., 2016). One set, a representative random sample drawn from across the country, is used to calculate the annual national moderator-teacher agreement rate, which is publicly reported.[36] The random selection of standards and students whose work is submitted allows valid inferences to be drawn regarding the reliability of internal assessment at a national level, and supports moderation fulfilling an accountability purpose (Hipkins et al., 2016). The second set comprises submissions of standards that have been deliberately selected from individual schools by NZQA staff, allowing moderation efforts to be targeted.[37] For the second set, NZQA requires schools to use a random sampling approach to select which students' work to submit (NZQA, 2013b), further reinforcing the accountability emphasis (Hipkins et al., 2016).

In contrast, at the time of data collection for this study, NZQA moderation of the tertiary sector contained no representative randomly selected set of submissions, and no public reporting of annual national moderator-teacher rates. Instead, the set of standards for submission, which comprises assessment towards NCEA and other qualifications, was selected from individual organisations by NZQA staff. NZQA made no specific requirements regarding how organisations select which student work to submit, instead asking for the

---

[35] Anecdotally, some TEOs do not require assessment results to be verified prior to being reported.
[36] This is called 'National Systems Check moderation' (NZQA, n.d.a).
[37] This is called 'School Check moderation' (NZQA, n.d.a).

purposeful selection of 'Achieved' (or higher) samples and only borderline 'Not Achieved' samples (NZQA, 2014c). While ongoing non-compliance with NZQA moderation could have ultimately resulted in the withdrawal of an organisation's consent to assess (NZQA, 2014c), consequences rarely appeared to have been imposed (Gerritsen, 2017a). The approach that NZQA has taken with the tertiary sector moderation results, at least in the past, belied a lower level of concern about quality control and public reporting for accountability with this sector than NZQA had with the secondary sector, suggesting, at the time of data collection for the present study, a lesser focus on maintaining public confidence.[38]

It seems likely that the apparent differences in NZQA's level of concern between the two sectors is reflective of differences in political interest and concern regarding the sectors. The Cabinet of the New Zealand Government approves the number of samples of student work from the secondary sector to be moderated by NZQA annually (Controller and Auditor-General, 2012), but does not appear to do so for NZQA's tertiary sector moderation, which strongly suggests that there is more political concern about ensuring the quality and credibility of NCEA qualifications in the secondary sector than tertiary sector qualifications. The Auditor-General conducted a performance audit report into NZQA's assurance of the internally assessed components of NCEA as delivered by the secondary sector (Controller and Auditor-General, 2012), but not into NZQA's moderation of the tertiary sector, also indicating that there has been less political focus on the quality and credibility of the qualifications awarded and student achievement in the tertiary sector than in the secondary.

2.4.3.3 Moderators must share common interpretations, understandings

For moderation to be effective for quality control and accountability purposes, moderators of the same assessment standards must share a common interpretation of those standards, and all moderators in an assessment system must share a common understanding of standards-based assessment and the role of moderators. In their qualitative study into external moderation that aimed to ensure inter-institutional comparability in the U.K. university sector, Bloxham et al. (2015) found widespread variation between external moderators in these aspects, which, they concluded, threatened the credibility of the system.[39] The moderators involved in the study ($n = 24$, from chemistry, history, nursing, and

---

[38] It is noted that in recent years, NZQA appears to have begun demanding more accountability from the tertiary sector, imposing sanctions and interventions on TEOs with ongoing non-compliance and poor performance in educational and assessment practices and moderation. (Refer Gerritsen, 2017a, 2017b, 2017c.)

[39] Although the sample was too small (~six per discipline) to enable reliable conclusions to be drawn.

psychology across 20 universities) were found to differ sufficiently within each discipline in their interpretation of assessment standards, the importance placed on different criteria, and the meaning assigned to the standards and criteria, to award manifestly different grades in 95% of cases. Furthermore, the researchers found differences in participants' understandings of the moderator's role. Informed by their findings, Bloxham et al. (2015) recommended that the role of external moderators be clearly documented, and regular calibration exercises for moderators be implemented. Bloxham et al. (2016) also found evidence to suggest that assessment standards were not effectively maintained by moderation. After analysing moderation practices in the Australian and U.K. higher education sectors, they concluded that not all of the observed moderation practices would effectively provide quality control. Instead, the researchers asserted that some moderation practices would only offer the *appearance* of assuring fair, accurate, and consistent assessment, and "provide largely a public image of systematic checking of standards" (2016, p. 648). In the course of conducting this literature review, no other research was found into the consistency or comparability of moderator judgements, interpretation of standards, or practice. As such, this highlights a gap in existing literature.

2.4.3.4 Empirical research: Impact of moderation on assessor judgements

Moderation processes and practices must be rigorous and robust to effectively ensure that assessments are sound and trustworthy, and to maintain legitimacy in the eyes of stakeholders. Two studies from the Australian higher education sector (Bird & Yucel, 2013, and O'Connell et al., 2016) sought to quantify the impact of moderation interventions on the accuracy and consistency of assessor judgements, both with positive results (described below). Apart from these, there appears to have been little recent empirical research into the effectiveness of moderation for quality control or accountability purposes, thus, representing another research gap.

Bird and Yucel (2013) collected data about an intervention in a first-year undergraduate biology paper in one university, where summative assessments (reports) were marked by eight casual assessors (refer 2.4.0.1). The intervention involved a one-hour assessor training workshop followed by a two-hour social moderation meeting (to calibrate assessors' judgement) prior to marking occurring. Bird and Yucel used a one-group pre-test-post-test quasi-experimental design (B. Johnson & Christensen, 2008) in their study, in which all

participants were pre-tested, subjected to the intervention, and then post-tested. In the pre-test phase, assessors each individually marked a set of 20 (cleaned) reports from previous years' cohorts, utilising normal practice, and awarding marks out of 100. In the post-test phase, assessors marked a second set of 20 (cleaned) reports (also from previous years' cohorts) in accordance with their current understanding and using the marking rubric introduced in the intervention. An expert assessor (an experienced assessor recruited from another university) also marked both sets of reports (as per pre- and post-tests). The expert assessor's marking was taken as accurate, and assessors' judgements were compared with that of the expert to determine accuracy. Assessors' judgements were compared with each other to calculate inter-rater reliability. The researchers found that the difference in marks awarded to the same report by individual assessors and the expert assessor reduced significantly between the pre- and post-tests. The differences between individual assessors did not decrease significantly. Thus, the intervention appeared to be effective at improving the accuracy of assessors' marking, but not at improving inter-assessor reliability. Certain aspects of Bird and Yucel's study design were stronger than others (stronger aspects included the use of a pre-test-post-test design, the number of reports used in the tests, and the use of an expert marker to provide 'accurate' judgements; weaker aspects included the lack of a control group, the small sample of assessors, and the lack of measures to ensure consistency in the expert marker's judgements). However, the findings of this study suggest that such interventions may be of assistance to improve the accuracy of assessors' judgements.

O'Connell et al. (2016) evaluated the effectiveness of the *Achievement matters: External peer review of accounting learning standards* project, a multi-institutional Australian intervention focused on nationally defined accounting threshold standards (refer 2.4.2). The researchers used a pre-test-post-test control-group experimental design (B. Johnson & Christensen, 2008), in which both groups were pretested, the treatment group only were subject to the intervention, and then both groups were post-tested. Participants (treatment: *n* = 30; control: *n* = 15) were lecturers who volunteered from 17 higher education institutions (15 universities, one community college, and one private provider). The pre-test comprised all participants individually assessing three samples of student work against two accounting standards, awarding marks from 0–100. The intervention involved a four- to six-hour-long face-to-face meeting in which social moderation was undertaken, involving discussion and debate about their own pre-test judgements, the standards, the assessment tasks, and new samples of student work. The post-test comprised participants of both groups individually reassessing the same samples of student work as were used in the pre-test. The judgements of the two groups were then compared. O'Connell et al. (2016) found that at the

pre-test stage, for both groups and both standards, there was sizable variation in assessor judgement. However, the reduction in standard deviation for the treatment group from pre-test to post-test was found to be significant for both standards. These findings indicated that, without intervention, there was a lack of comparability in assessor judgements for the standards between institutions, and that the intervention effectively reduced variability in judgements (and therefore increased comparability between assessors). While the sample sizes (participants and items of student work used in the pre-test and post-test) were small, and the groups were not randomised, the inferences of the findings are credible because of the relatively large number of institutions involved, the relative comparability between the two groups, and the strength of the experimental design.

To date, the calls of Bloxham et al. (2016) and S. Johnson (2013) for empirical research into the effectiveness of moderation have barely been heeded. The findings of Thomas et al. (2014) and Ward and Thomas (2016) emphasise a need for research; the lack of reliability found by both studies strongly suggest that effective moderation for quality control is needed in the New Zealand context.

2.4.3.5 Research using perception-based or self-report data

In contrast to the dearth of quantitative empirical studies, many studies into moderation for accountability-type purposes that used perception-based, self-report data have been published, including research into stakeholders' perceptions of the credibility of the standards or qualifications involved.

Several studies in the vocational education and training sector in Australia were prompted by a lack of stakeholder confidence in the quality of training and assessment. Halliday-Wynes and Misko (2013) reported that concerns about whether graduates always had the competencies and work-readiness indicated by their qualifications led to extensive policy debates about risks to the credibility of qualifications offered in that sector. Halliday-Wynes and Misko's study examining stakeholder perceptions of assessment practices within parts of the sector identified by the national regulator as being of high risk, and the subsequent study of Misko et al. (2014), were initiated in response to these debates. Halliday-Wynes and Misko (2013), in a small-scale qualitative study, asked assessors in public and privately-owned Registered Training Organisations about their assessment practices. Their findings suggested that a lack of assessment competence and expertise among assessors posed a risk to the perceived trustworthiness of assessments and qualifications, and that the risk was

exacerbated by the widespread practice of organisations developing or modifying assessment instruments for their individual contexts (despite their lack of assessment expertise). Furthermore, Halliday-Wynes and Misko found substantial cross-sector variability in moderation approach, focus, and implementation frequency. The researchers concluded that the lack of systematic internal and external moderation processes reduced stakeholders' confidence in the accuracy and comparability of assessments (and therefore, qualifications awarded), in the sector.

New Zealand stakeholder perception data pertaining to NCEA moderation and support for NCEA have been collected by the NZCER secondary school surveys (refer 2.2.2). While the last three surveys collected perception data regarding NCEA moderation from teachers and principals, moderation was framed as being for improvement or professional learning in most survey items (see Hipkins, 2013; Wylie & Bonne, 2016). However, in both 2009 and 2012, principals and teachers were asked if moderation took too much time (which related to compliance with requirements, and was, therefore, relevant to quality control purposes). In each survey, approximately two thirds of teachers and principals felt that moderation took too much time; those teachers with low morale, or who were middle managers (e.g., Head of Department), were more likely to feel that way (Hipkins, 2013). In the 2015 survey, a different compliance-type item was asked of teachers regarding the clarity of NZQA moderation requirements.[40] Almost half of teachers agreed that the requirements were clear, but almost one third disagreed that they were (Wylie & Bonne, 2016). Data about stakeholder support for NCEA, and its credibility, provide evidence relevant to the accountability and maintenance of public confidence functions of moderation, given that internal assessment is a central feature of NCEA (Hipkins et al., 2016). The surveys have gauged the level of support for NCEA from key stakeholder groups[41] since its inception. They have revealed a stable or slightly increasing level of support for NCEA among most groups surveyed (Wylie & Bonne, 2016): Principals have been almost unanimous in their support since 2009; two thirds of teachers have supported NCEA in each survey; support among parents has increased from one third in 2006 to remain steady at just over half since 2012; and trustees showed increasing levels of support until 2012, (although this decreased from three quarters to two thirds in 2015). Since 2006, the surveys have also asked respondents whether they perceived that the wider community saw NCEA as a credible qualification (Hipkins, 2013; Wylie & Bonne, 2016). The proportions who agreed that the

---

[40] Principals do not appear to have been asked any compliance-type items about moderation in the 2015 survey. Refer Wylie and Bonne (2016).
[41] Parents and whānau, teachers, principals, and members of Boards of Trustees (Wylie & Bonne, 2016).

NCEA was seen as credible increased from 2006 to 2015 (from half to almost all principals; and from one third to two thirds of teachers, trustees, and parents). No equivalent research into New Zealand tertiary-sector stakeholder perceptions of NZQA moderation, or qualifications including NCEA, appears to have been conducted. As such, this is another gap in the literature.

Two studies addressed the moderation of National Standards in New Zealand primary and intermediate schools. The first (Wylie & Bonne, 2014) reported on the findings of the 2013 NZCER national survey of primary and intermediate schools. The survey was administered to key stakeholders from a representative sample of primary and intermediate schools, with resulting sample comprising *n* (principals) = 180, *n* (teachers) = 713, *n* (trustees) = 277, and *n* (parents) = 684. The implementation of moderation for National Standards was a requirement (Ministry of Education, 2010), so survey results relating to the implementation of moderation processes in schools has relevance to quality control functions.[42] Teachers reported a widespread increase in moderation conducted by schools in response to the introduction of National Standards: 70% reported that there had been an increase in moderation between teachers of the same year level, and almost 60% reported an increase in moderation between teachers of different year levels (Wylie & Bonne, 2014). The survey also sought to ascertain stakeholders' confidence in the achievement data provided by National Standards (Wylie & Bonne, 2014). Principals' and teachers' perceptions about the clarity of the standards in terms of expected achievement, ease of making reliable judgements against them, or how robust the standards were (7-30% of principals, and 15-49% of teachers responding in the positive) suggested that neither group was particularly confident of judgements at their own schools. In addition, confidence in the reliability of National Standard data was found to be very low: Only 5% of principals, 10% of teachers, and 16% of trustees agreed that current National Standards data from schools provided a reliable picture of student performance. In contrast, 87% of principals, 70% of teachers, and 51% of trustees *disagreed* that the data provided a reliable picture.[43]

The study by Ward and Thomas (2016; refer 2.2.3.1) also considered moderation of New Zealand's National Standards. The researchers collected self-report survey data from principals (*n* [2014] = 76) and drew on previously collected data from the *National Standards: School sample monitoring and evaluation project* to identify trends from 2010 onwards. In 2013 and 2014, almost all principals were confident in the consistency of

---

[42] The survey items that framed moderation in terms of improvement and professional learning are discussed in section 2.4.4.2.
[43] Parents were not asked about their perceptions of the reliability of the data.

assessor judgements made within their own schools, but three quarters lacked confidence in the between-school consistency of judgements. The part of Ward and Thomas' (2016) study that was based on survey data relating to moderation practices had several limitations, thus, some of their conclusions must be treated with caution. One limitation was that the survey items framed moderation only as social moderation (e.g., in terms of teachers meeting to discuss and moderate judgements), meaning that data about other moderation approaches used may not have been captured. As a result, the number of schools implementing moderation may have been under-represented. Notwithstanding this limitation, in 2014, for reading and mathematics standards, 14% and 17% of schools respectively were reported as not having moderated the standards that year. Further, while the percentage of schools reported as implementing moderation of the reading and writing standards via 'systematic discussions' gradually increased year on year (to approximately 60% for reading and 80% for writing in 2014), the percentage reported as moderating the mathematics standards via this approach varied dramatically from year to year (46% in 2010, 90% in 2011, 54% in 2014). Another limitation was the narrow definition of 'effective' moderation[44] applied. Effectiveness of moderation was defined only in terms of the sampling approach used (the specific approach that was defined as effective was where samples were selected from above and below grade boundaries). No consideration was given to the moderation process itself, its frequency or timing, or the uses made of moderation decisions. Such a narrow definition limits the validity of any inferences drawn regarding the effectiveness of moderation occurring within schools: Not only are these other components integral to the effectiveness of moderation, the purposes for which moderation results are used will likely impact on how moderation samples are selected. For example, an approach in which samples are selected from either side of grade boundaries would be appropriate for improvement purposes, where using random selection would be appropriate for accountability purposes. Furthermore, there would likely be more than one sampling approach that would be appropriate for a purpose: For instance, selecting those judgements for which there was inconsistent assessment evidence would seem appropriate for quality control and accountability purposes, as would moderating *all* assessor judgements. However, by applying such a narrow definition, Ward and Thomas determined that only one third (for writing standards) to one half (for reading and mathematics standards) had used effective moderation processes, which seems likely to be a flawed inference.

---

[44] The definition of 'efficient' moderation that was applied was similarly narrow and flawed. As such, the inferences about the proportion of schools using efficient moderation processes in 2014 should also be treated with caution.

2.4.3.6 Potential for adverse consequences

Potentially harmful, often unintended, consequences of quality assurance requirements when part of high-stakes accountability systems have been widely reported. These include negative impacts on teaching, student learning, and curriculum delivery (Altrichter & Kemethofer, 2015; Ehren & Hatch, 2013; Ehren et al., 2015; Gustafsson et al., 2015), a reduced focus on quality improvement (NZQA, 2012a), an increased focus on making superficial changes to result in short-term improvements against set expectations at the expense of more meaningful changes and improvements (e.g., Ehren et al., 2015; Misko, 2015a), presenting an orchestrated, overly-positive image to external inspectors (e.g., Ehren et al., 2015), and adverse impacts on the workload of teachers and administrators (e.g., Ingvarson et al., 2005; Wong & Li, 2010; Wylie, 2013). Reported adverse consequences of moderation with a strong accountability function include teachers being less inclined to implement creative or integrated approaches to internal assessment (Hipkins et al., 2016), the adoption of risk-averse marking approaches (Grainger et al., 2016), and only submitting samples for moderation in which teachers are confident in the grades awarded—thereby reducing opportunities for professional learning (Hipkins et al., 2016). The potential for moderation data to be used for an accountability function was observed to be a barrier for some organisations to fully engage in a cross-sector, inter-school social moderation initiative and fully utilise the improvement and learning opportunities available (Grant, 2012; discussed further in section 2.4.4).

2.4.3.7 Moderation for accountability purposes: Summary

To summarise, moderation is widely used for accountability and quality control purposes, as a process through which to ensure the credibility and integrity of internal assessments and assessment practices, and thus to maintain public and stakeholder confidence in assessment standards and qualifications awarded. Differences in the moderation regimes that apply to New Zealand's secondary schools, as compared with TEOs, illustrate the extent to which political desire to maintain public confidence in the education and national qualifications systems determines the requirements of such regimes: At the time of data collection, NZQA's moderation approach for the secondary sector was more robust and stringent than the approach taken for the tertiary sector. There is a relative paucity of empirical evidence of the effectiveness of moderation to improve accuracy and consistency of assessor judgements, and as such, more empirical research is required. In contrast, there is a sizable body of evidence of stakeholders' perceptions regarding moderation for

accountability and quality control. This evidence affirms the need for effective moderation to fulfil these purposes, to increase public confidence in assessment standards, education, and qualifications awarded.

2.4.4 Moderation for improvement and learning purposes

Moderation and other quality assurance processes can also be used to assist organisations or individuals to learn and improve their actions, practices, or performance (Danø & Stensaker, 2007; Ehren et al., 2013; Misko, 2015a). An organisation itself can learn through individuals within the organisation, by those individuals' learnings being adopted and embedded in the organisation's processes and practices (Millward & Timperley, 2010). Moderation can be used to help improve the quality of assessment instruments, processes and practices, or assessor judgements, and to provide professional learning opportunities, including about standard requirements, assessment or teaching practice, or curriculum matters (Bloxham et al., 2016; Hipkins & Robertson, 2011; Klenowski & Wyatt-Smith, 2014). Thus, another set of closely-related purposes for which moderation and other quality assurance processes are used involve improvement and learning: quality improvement, and providing opportunities for professional or organisational learning, which can occur at system-, organisation-, department-, or individual- level (Adie, 2012; Grant, 2012). These purposes feature in moderation and quality assurance regimes across all levels of education, from early childhood (Wong & Li, 2010) to university (Adie et al., 2013). The rest of this section examines research findings regarding the effectiveness of moderation for learning and improvement, the mechanisms through which such functions occur, stakeholders' perspectives of moderation for learning (from a New Zealand context), and factors that influence whether people avail themselves of the learning opportunities present. Initially, the eminence of learning and improvement purposes in moderation regimes is discussed.

The primacy of improvement and learning as intended purposes of moderation and other quality assurance regimes varies. In some, improvement and learning are the prime drivers—either instead of, or alongside, accountability. For example, in the assessment and moderation system implemented in New Zealand's workplace training sector by the Building and Construction Industry Training Organisation, moderators were considered a part of the assessment team, providing assessors with support and advice (Vaughan et al., 2012; 2.4.0.1). Further, national moderation meetings were focused on providing professional development opportunities for all involved through open conversations, and a shared purpose of improving training and assessment practice. In other regimes, improvement and

learning are considered to be consequential to the primary drivers of accountability and quality control—as a secondary intention, or simply as a result (Law, 2010). For example, NZQA moderation appears to place less weight on improvement and learning as purposes than on accountability and quality control.[45] Schools are required to act on advice from NZQA moderation (NZQA, n.d.a), while TEOs are recommended to use NZQA's moderation feedback to inform assessment practice and design, and internal moderation processes (NZQA, 2014c). However, as Hipkins et al. (2016) observed, there is little in the NZQA moderation process that guarantees professional learning or improvement.

2.4.4.1 Research into moderation for improvement and learning purposes

Numerous studies conducted across the education sector in recent years have examined moderation with learning and improvement functions. Some were based on empirical and self-report data, and others on self-report data only. A number of these studies have already been introduced: Adie et al. (2012; 2.2.1), Bird and Yucel (2013; 2.4.0.1 and 2.4.3.4), Black et al. (2010, 2011; 2.2.3.1), Crimmins et al. (2016; 2.4.0.1, 2.4.1), Hipkins (2013; the 2012 NZCER survey—2.2.2), O'Connell et al. (2016; 2.4.3.4), and Wylie and Bonne (2014; 2.4.3.5). Studies from the compulsory education sector (Adie, 2012; Connolly, Klenowski, & Wyatt-Smith, 2012; Controller and Auditor-General, 2012; Grant, 2012), and the Australian university sector (Beutel et al., 2017; Watty et al., 2014) are now introduced, before findings regarding moderation for improvement and learning purposes are discussed.

Adie's (2012) longitudinal study focused on online social moderation. Adie analysed a subset of data from the Australian Research Council Linkage project (involving middle school teachers in Queensland; refer 2.2.1) pertaining to the online moderation option of the project. This involved synchronous meetings via an online-meeting technology platform, in which grades for student work samples were moderated. The sample comprised 50 teachers (from three disciplines, different year groups, and diverse localities and sociocultural contexts). Data collection via observations of online moderation meetings, and pre- and post-meeting interviews occurred over two years, and follow-up interviews were conducted in the third

---

[45] Hipkins and her colleagues charted a shift in the emphasis of NZQA moderation from improvement and learning in the early years of NCEA, to accountability, driven by public and political demand. They mused that "to achieve public acceptance of the NCEA system [with its internal assessment component], a moderation system that, at the outset, entailed a balance of accountability and professional learning would have been wise" (Hipkins et al., 2016, p. 133).

year. The combination of self-report and observation data enabled triangulation, and the follow-up interviews enabled insights into how participants had subsequently enacted their learning in their own contexts, post-project. Connolly et al. (2012) also analysed a subset of data from the Australian Research Council Linkage project. These data were transcripts from interviews with a sample of 67 teachers from 24 schools, conducted prior to and after face-to-face and online moderation meetings, allowing comparison of participants' perspectives. The previously-noted limitations of the Linkage project (participants were new to standards-based assessment and moderation—refer 2.2.1) also apply to Adie's study and that of Connolly et al.

New Zealand's Controller and Auditor-General (2012) conducted a performance audit of NZQA and, as a part of her audit, collected perception data via survey from teachers about NZQA moderation. A census approach was used to obtain the survey sample, in which all secondary schools were asked to distribute the survey link to all teachers. Of the resulting sample (n = 1,780), 1,500 respondents indicated that a submission of theirs had been moderated by NZQA in the preceding two years. The survey items that directly pertained to NZQA moderation focused on opportunities for improvement and learning. While this study collected self-report perception data only, the resulting sample was large (with most respondents having had recent experience of NZQA moderation), lending weight to the survey results.

Grant (2012) reported on a four-year initiative involving staff from all primary and secondary schools in the jurisdiction of one Scottish educational authority. The initiative involved recurring cross-sectoral social moderation events and aimed to provide professional development for teachers, develop sustainable approaches to assessment within the schools, and develop an inter-sectoral community of practice that was sustainable. Annually, the initiative comprised moderator training to calibrate teachers' moderation judgement, before a social moderation meeting in which groups of teachers moderated marked samples of student work against assessment criteria. Each group contained primary and secondary teachers from the same cluster of schools. A total of 518 participants (from an undisclosed number of schools) were involved in the initiative over the four years: 432 from primary schools and 72 from secondary schools (each comprising 90% teachers, 10% management staff), and 14 educational authority staff. Focus group, participant evaluation, moderation discussion, participant log, and moderation meeting observation data were collected and qualitatively analysed. Grant (2012) did not report on the reliability of assessor judgements or teacher-moderator agreement rates but focused on the potential of cross-sectoral social moderation in developing sustainable assessment cultures. Therefore, the impact of the

initiative on the reliability or accuracy of assessment is unknown. The sample size, multiple data sources, and longitudinal nature of the study provide credibility to Grant's findings.

In their qualitative study, Beutel et al. (2017), identified the moderation practices utilised in one faculty within a large urban Australian university, to identify challenges to moderation that supported teaching and student learning. The researchers interviewed a purposive sample of 25 academic staff (mostly unit coordinators; the balance, tutors), selected from a broad range of courses in terms of size, configuration, and delivery approach.[46] A limitation of the study is that only self-report perception data were analysed, precluding the possibility of corroborating their findings. Despite this limitation, the study's findings provide insight into moderation practices in a university context.

Watty et al. (2014) gathered participant perception data from an inter-institutional social moderation initiative that had aimed to develop a shared understanding among academics from ten Australian universities of nationally set accounting threshold standards. (This was an early phase of the *Achievement matters: External peer review of accounting learning standards* project that O'Connell et al., 2016, studied; refer 2.4.3.4) The initiative that Watty et al. reported on involved more than 30 academics who participated in three cycles of a three-stage peer review and calibration process. Each cycle reflected a one-group version of the pre-test-post-test design as used by O'Connell et al. in which all participants received the treatment (B. Johnson & Christensen, 2008). Although no control group was included in the study design, the quasi-experimental design and recurrent nature of the multiple cycles (providing a longitudinal element) mean that the findings of this study provide valuable and credible insight into participants' perceptions. That no quantitative analysis appears to have been conducted on the impact on reliability of assessor judgements through this study was a missed opportunity.

2.4.4.2 Moderation can result in learning and improvement

Multiple studies have provided evidence of moderation effectively resulting in learning and improvement, including two that have already been discussed. The finding by O'Connell et al. (2016) of a social moderation workshop reducing inter-rater variability (see 2.4.3.4) indicated that social moderation enhanced the degree of calibration between assessors' judgements. Bird and Yucel's (2013) finding that a social moderation session improved the

---

[46] The same research team had analysed the same data set for a previous study—Adie et al. (2013)—which will be discussed in 2.4.6.

accuracy of marking also indicated a certain effectiveness (although they found that the intervention made no real difference to the reliability between assessors). Black et al. found that participants took learning from each phase of moderation that occurred in their longitudinal intervention, from developing greater assessment capacity and literacy through their critique, redevelopment, and refinement of assessment instruments, to the learning that participants took "as they established communal standards through grappling with differences between their own judgements and those of their colleagues" (Black et al., 2010, p. 225). Adie (2012), Adie et al. (2012), and Grant (2012), also found that moderation could result in learning and improvement.[47]

2.4.4.2.1 Participants recognise opportunities afforded

The learning and improvement functions of moderation appear to have also been recognised by participants in various studies. Participants in the studies of Black et al. (2010, 2011), Crimmins et al. (2016), and Watty et al. (2014), reported that the moderation initiatives provided them with valuable professional learning and development. Wylie and Bonne (2014) found that teachers who participated in the 2013 NZCER primary and intermediate schools survey saw moderation as having had the most positive impact on their teaching practice of any of the changes to their school's assessment practices that had resulted from National Standards, and 70% felt that moderation of their judgements with other teachers had provided useful insights into their own practice. Principals concurred: Most saw moderation as the most useful addition at a school-level of everything associated with National Standards, and over 80% considered the moderation of assessor judgements to provide valuable professional learning.

2.4.4.2.2 Opportunities can relate to teaching, student learning, and assessment cycle

It has been found that the learning that results from moderation can relate to any aspect of the teaching, student learning, and assessment cycle: for example, to assessment standards, criteria, instruments, practice, or judgement-making processes, required qualities

---

[47] Each of these studies involved social moderation as a key component in the intervention. However, considering these positive findings, it is unclear how effective or feasible social moderation would be at scale (i.e., involving thousands of geographically dispersed assessors). This issue is returned to in 2.4.5.1.

of student work, curriculum matters, pedagogy, teaching practices and resources, and student learning (Adie, 2012; Adie et al., 2012; Black et al., 2010, 2011; Grant, 2012; Hipkins & Robertson, 2011; Mottier Lopez & Pasquini, 2017; Wyatt-Smith et al., 2010). Furthermore, participants in various studies have reported that their learning through moderation had impacted positively on their teaching or assessment practice (e.g., Black et al., 2011; Crimmins et al., 2016; Grant, 2012; Hipkins, 2013; Watty et al., 2014; Wylie & Bonne, 2014).

2.4.4.2.3 Scheduling of moderation activities

There is some suggestion that the scheduling of moderation activities can impact on learning opportunities and on the potential impact of learning. Bloxham et al. (2016) recommended that moderation activities which focused on assessment instruments, requirements of standards, and calibration of assessor judgements, be implemented before teaching occurred, to enable the learning that teachers took from the moderation to inform their teaching practice. This was corroborated by the finding of Beutel et al. (2017) that moderation practices that were implemented throughout the teaching and assessment period were perceived to allow people's developing understandings to inform their teaching practice.

2.4.4.2.4 Development of shared understandings

Shared understandings were found to be progressively developed among participants through discussion and negotiation in moderation contexts, including of the requirements and terminology of the standards and criteria, the qualities of student work that reflected different grades, and judgement-making processes (Adie, 2012; Adie et al., 2012; Black et al., 2010). Participants were also found to have developed a shared language with which to communicate about standards and assessment matters (Adie et al., 2012; Black et al., 2010). Beutel et al. (2017) found that strategically scheduled moderation practices that were implemented throughout the teaching and assessment period, and that involved the teaching team in collaborative negotiation, were seen by their study participants to assist the development of shared understandings of standards and assessment. Watty et al. (2014) found that their study participants recognised that the assessment standards they were using were complex, but believed that the social moderation process involved in their inter-university calibration initiative helped to develop a shared understanding of those standards among multiple assessors. Connolly et al. (2012) also found that participants considered that

social moderation assisted teachers to develop a common understanding of the standards, and thus enable fair and consistent marking.

Bird and Yucel's (2013) study was one in which the intervention of interest did not effectively enable participants to develop a shared understanding of the standard: The intervention made no real difference to the consistency of judgements between assessors. Bird and Yucel concluded that, given the assessors' different levels of experience, the intervention had likely been too short[48] to achieve inter-rater reliability. The lack of opportunities that the intervention provided for assessors to develop a common understanding of the standard, (e.g., through scheduled social moderation of student evidence) was also a likely contributor to this result. It takes time to develop a shared understanding of assessment criteria and the quality of student work that meets them (Klenowski & Wyatt-Smith, 2014); a short training workshop and social moderation meeting, both of which occurred pre-marking, did not achieve sufficient calibration of assessors.

2.4.4.2.5 Impact on teachers' professional confidence as assessors

Moderation initiatives with a focus on learning and improvement can increase teachers' confidence in themselves as assessors. Over the course of their intervention, Black et al. (2010) observed participants grow in confidence in their own assessment knowledge. The participants corroborated this observation, reporting that their own confidence in their knowledge and judgement associated with all parts of the assessment cycle improved through the intervention (Black et al., 2011). Connolly et al. (2012) found that an outcome of social moderation that participants specifically valued was an increase in teachers' confidence as assessors. O'Connell et al. (2016) collected self-report data from participants in their treatment group, via pre- and post-treatment surveys. Parametric tests were used inappropriately with ordinal survey data, meaning that the inferences drawn must be treated with caution. However, participants' confidence in their own assessor judgements against the accounting standards appeared to have increased after the treatment. Participants in the studies of Crimmins et al. (2016), and Watty et al. (2014), reported that through each, their professional confidence as assessors increased.

---

[48] The intervention occurred prior to assessors undertaking marking, and comprised a one-hour assessor training workshop, followed by a two-hour social moderation meeting (see 2.4.3.4).

2.4.4.2.6 Learning is individualised

The learning that participants took from moderation events was found to be individualised (Adie, 2012; Adie et al., 2012). Adie et al. (2012) found that participants' development in understanding in moderation contexts was influenced by diverse factors beyond the explicitly-stated criteria in the standards, including how the standards were represented in qualitative and quantitative terms, and participants' own knowledge base, prior experience, and internalised notions of the standard. Adie (2012) observed that, while shared meanings were negotiated within social moderation contexts, individual participants had their own interpretations of those meanings. Participants took their own learning from the experience, which sometimes differed from the learning taken by others from the same experience. Furthermore, individual participants later enacted their learnings and interpretations of those shared meanings into their own contexts in different ways.

2.4.4.2.7 Role of collegial discussion and debate

The dialogue and negotiation that occurs through moderation (e.g., in social moderation meetings), have been found to provide learning opportunities (Adie, 2012; Adie et al., 2012; Black et al., 2010; Grant, 2012), reflecting the findings of previous research (as reported by Hipkins and Robertson, 2011, in their review of literature). Furthermore, study participants recognised this. Those moderation practices involving staff in collegial discussion and debate were perceived by Beutel and colleagues' (2017) participants to provide opportunities for learning, and those practices that did not were seen to hinder learning or any development of shared understandings. The casual tutors in the study of Crimmins et al. (2016) felt that the pre-marking calibration meetings, and opportunities for collegial dialogue that these contained, provided highly valuable professional learning. Participants in Grant's (2012) study reported that they valued the collegial conversations that occurred in the moderation meetings involved in the study. Likewise, participants in Bird and Yucel's (2013) study reported that they valued the social moderation meeting in the initiative, and particularly the opportunity for collegial conversation.

2.4.4.2.8 Role of moderation feedback

Receiving moderation feedback can provide opportunities for professional learning. Adie et al. (2012) observed that receiving feedback in a social moderation context prompted self-reflection on the part of recipients, and participants in Crimmins and colleagues' (2016) study

identified the feedback on their marking that they received individually from the 'expert' moderator as providing learning opportunities. The content, focus, clarity, and tone of moderation feedback have all been found to impact on the resulting learning opportunities for recipients (Bloxham et al., 2016; Colbert, Wyatt-Smith, & Klenowski, 2012). Recipients' reflective engagement with that feedback was also observed to impact on their learning opportunities (Adie et al., 2012). The learning opportunities appeared to be enhanced when there were opportunities for the recipients to seek clarification and engage in dialogue with the moderator(s) regarding that feedback (Adie et al., 2012; Crimmins et al., 2016). However, Adie (2012) found that how recipients reacted to feedback (e.g., whether they experienced it as affirming or threatening), impacted on their levels of engagement and learning. Adie observed that some participants appeared to feel threatened by having their judgements challenged, and subsequently withdrew or disengaged from the moderation, while others appeared to find the experience of having their judgements endorsed by colleagues legitimising or affirming, and this seemed to promote their engagement and learning. Adie (2012) concluded that participating in moderation would not automatically result in learning.

2.4.4.2.8.1 Stakeholder perspectives of NZQA moderation feedback

In the New Zealand secondary school context, stakeholders' perspectives on NZQA moderation feedback have been canvassed. The Controller and Auditor-General's survey items asked about the content of the moderation reports (in terms of usefulness of feedback about assessment materials and assessor judgements), and the clarity and tone of that feedback. Approximately 80% of teachers surveyed agreed that the feedback was useful, clear, and professional and supportive in tone (Controller and Auditor-General, 2012). In the 2012 NZCER secondary school survey, most items relating to moderation framed it as having improvement and learning purposes and focused on moderation feedback. The items addressed using moderation feedback to inform revision of assessment instruments, the helpfulness of NZQA moderation reports, the teaching team's agreement with NZQA moderation feedback, and the helpfulness of NZQA feedback for clarifying the intention of the assessment standards. Hipkins (2013) reported that one half to three quarters of teachers responded positively to each item (i.e., agreed that they found NZQA moderation reports helpful, etcetera). Teachers with lower levels of morale, or who taught mathematics, science, or computing, were less likely to view NZQA moderation as having learning and improvement functions. Conversely, teachers with higher levels of morale, or who taught social sciences, arts, or commerce, were more likely to. Furthermore, teachers who

indicated that they had stronger professional learning networks were more likely to view NZQA moderation as having learning and improvement functions, and vice versa (Hipkins, 2013). Principals were asked whether NZQA moderation provided teachers with valuable insights into the expected levels of achievement for assessment standards, and a large majority (86%) agreed that it did (Hipkins, 2013). Even though principals and teachers were asked slightly different things by the NZCER surveys, the results suggested that in 2012, principals saw NZQA moderation as having a somewhat stronger learning and improvement function than teachers did.

The Controller and Auditor-General's (2012) survey results suggested teachers found NZQA moderation feedback to be more helpful than the 2012 NZCER secondary school survey results did (Hipkins, 2013). There are multiple possible explanations for the differences in results (approximately 80% versus 50%-75%). These include possible differences in the samples: The NZCER sample was drawn from the relevant teachers' union database (Hipkins, 2013), while a quasi-census approach was used by the Controller and Auditor-General; or that teachers responded differently due to who was conducting the survey (Controller and Auditor-General versus NZCER). Another explanation could be the timing of data collection for each survey: The Controller and Auditor-General conducted her survey in November 2011 (Controller and Auditor-General, 2012), where the NZCER survey was conducted in the third school term of 2012 (Hipkins, 2013). Further possible explanations include differences in NZQA's moderation reports from 2010/2011 to 2012 that some teachers in Hipkins' (2013) sample reacted to; or that in 2012 teachers were moderated on new achievement standards (part of the review of New Zealand Curriculum-based assessment standards; NZQA, 2012c), and this moderation was more problematic. It seems plausible that any of these possibilities may have contributed to the differences in findings.

2.4.4.2.9 Interpersonal factors influence engagement in social moderation

Previous research has established that various factors influence people's willingness to fully participate in social moderation, and, therefore, impact on the learning opportunities present in the situation for them. These include the level of trust within the group, how comfortable and safe group members feel with each other, and the extent to which they each feel that the content and process of the moderation event have relevance for them (Hipkins & Robertson, 2011). Because many of the learning opportunities reside in the social interactions and

situation, the group dynamics and power relationships present influence the extent to which social moderation situations provide meaningful learning opportunities for participants (Bloxham et al., 2016). Grant (2012) found that it was important for participants in a social moderation context to get to know each other, and build relationships and trust within the group, to fully engage in the process. She observed that to begin with, participants in her study were reluctant to share their opinions and expose themselves to critique. Participants' initial reluctance appeared to be exacerbated by their lack of knowledge about the expertise of teachers from the other sector, and their subsequent assumptions about other teachers' expertise compared to their own (e.g., curriculum and subject knowledge). However, she observed this reluctance abate as a groundwork of trust and relationships within the group was built. The relationship-building opportunities provided by social moderation events appear to be valued by casual tutors in a university context. Participants in Bird and Yucel's (2013), and Crimmins and colleagues' (2016), studies reported that they valued the opportunity to develop a sense of connection, community, belonging, and collegiality with their colleagues.

2.4.4.2.10 Communities of practice

Scholars including Hipkins and Robertson (2011), and Klenowski and Wyatt-Smith (2014), have posited that social moderation approaches can function as communities of practice, when appropriate characteristics are present (i.e., relationships are built between members, members are engaged in the group's enterprise, there is a shared purpose within the group, and there are shared resources). However, social moderation situations will not *automatically* function as communities of practice; instead "appropriate focus, flow, and facilitation of [the group's] shared conversations and individual reflective learning" are required for them to do so (Hipkins & Robertson, 2011, p. 18).

Grant (2012) reported that teachers from clusters of schools were intentionally grouped to encourage the development of communities of practice; and observed that these communities did develop. In addition, she found that the initiative resulted in development opportunities for a wider cohort: Participants' learning impacted on their own practice and on the practice of their colleagues. Furthermore, Grant (2012) found that, through the development (outside of the research context) of cross-sector clusters of teachers, within-cluster inter-school learning and development occurred. Watty et al.'s (2014) finding that their study participants intended to disseminate their learnings into their own communities of practice also reflects the opportunities for wider development that moderation provides.

In the context of external moderation of GCSE course work (which uses an expert moderation approach), Crisp (2017) conceptualised external moderators as existing in a community of practice, albeit one in which there were periods of more-intense shared activity interspersed with lulls, in line with the assessment and moderation cycle. Crisp drew on her previous studies into the marking of internal assessment (Crisp, 2012, 2013), and proposed a conceptual model of judgements for internal assessment. This model comprised an "inner frame" of the assessment event itself (and the assessor judgements made by teachers), sitting within an "outer frame" of the social context within which high-stakes internal assessment occurred. Crisp's model conceptualised the external moderators as belonging to a "central" community of practice, and teachers to an "extended, dissipated" community of practice (Crisp, 2017, p. 34); the understandings and norms of the central community were communicated to the extended community via published resources and guidance documents, training, and moderation feedback. Many components of Crisp's (2017) model are evident in the high-stakes internal assessment that occurs in New Zealand's secondary and tertiary sectors, suggesting that communities of practice exist in this context.

2.4.4.3 Moderation for improvement and learning purposes: Summary

In summary, there is a growing body of evidence showing that moderation can provide learning and improvement opportunities to those involved, although these opportunities do not always result in learning or improvement actually occurring. The opportunities can relate to any aspect of the teaching, student learning, and assessment cycle, and the resulting learning (on the part of teachers) is individualised. Much of the evidence pertains to social moderation approaches, where the opportunities arise in part through the dialogue that occurs in these contexts, and through which shared understandings are negotiated. Learning and improvement opportunities have also been found to arise through receiving moderation feedback, although the content, tone, and timing of that feedback appears to impact on those opportunities. Evidence suggests that participants often recognise and value that moderation provides learning and improvement opportunities. Furthermore, many participants experience an increase in confidence as assessors as a result of this learning.

2.4.5 Can moderation have both accountability and improvement purposes?

There is some contention as to whether moderation, or other quality assurance processes, can fulfil both accountability/quality control and improvement/learning purposes. Scholars including Danø and Stensaker (2007), Ratcliff (2003), and Van Kemenade et al. (2008), have argued that there is often tension between accountability and improvement purposes in quality assurance regimes. Beutel and colleagues' (2017; 2.4.4.1) study provided a moderation example of this: Some of the moderation practices used within the university were found to focus on meeting the accountability requirements, but not on developing shared understandings between staff of the relevant academic standards, and so failed to equip staff to make consistent judgements.

In contrast, Harvey and Newton (2007) argued that tension between accountability and improvement purposes was unnecessary. They conceptualised the two purposes as distinct, and only partially related, dimensions (not as opposing ends of the same continuum). Quality assurance processes that resulted in greater compliance with requirements sometimes (but not always) resulted in improvements in quality, and quality assurance processes that resulted in quality improvements sometimes (but not always) resulted in greater compliance. As Harvey and Newton argued, improvement and learning purposes function alongside accountability and quality control in many quality assurance and moderation contexts. Moderation or other quality assurance processes with an accountability or control focus may result in improvements and opportunities for learning. For example, in the primary school sector in New York City, Ehren and Hatch (2013) found that most quality improvement resulted from schools' preparation for external inspection visits, and from acting on the inspection feedback received; Wong and Li (2010; 2.3.0.2.2) found a similar thing in kindergartens in Hong Kong. Crisp's (2017; 2.4.4.2.10) conceptual model of internal assessment judgements also reflected this duality. In her model, interpretations, understandings, and practices from the external moderator community of practice were fed into, and informed, the teachers' community of practice. Crisp's model also suggested that internal moderation using a social approach facilitated teachers' learning and internalising of those understandings and practices as communicated from the external moderators (Crisp, 2017). Thus, moderation with an improvement and learning focus might result in greater comparability of assessment, and as such, satisfy accountability and control requirements. The Australian multi-university initiative reported on by Watty et al. (2014; 2.4.4.1), and O'Connell et al. (2016; 2.4.3.4), provided an example of this.

2.4.5.1 More than one moderation approach required

For moderation systems to effectively fulfil accountability and improvement purposes, it appears that they need to contain more than one approach, with strategic scheduling of moderation events. For example, Crimmins et al. (2016; refer 2.4.0.1) found that through implementing a moderation process that incorporated several different moderation approaches, participants perceived that the functions of accountability/quality control, and learning and improvement, were met. As shown earlier, expert moderation approaches are widely thought of as effective for accountability purposes. However, in themselves, expert moderation approaches may not be particularly effective for improvement purposes, or at providing opportunities for professional learning and the calibration of assessors' judgement (Bloxham et al., 2016; Hipkins et al., 2016). On the other hand, social moderation approaches are likely to be effective at developing shared understandings among participants and assisting to enhance comparability of judgement within those communities of practice. However, as Sadler (2012) observed, social moderation tends to be localised in scope. Social moderation approaches through which participants collaborate and negotiate meaning, when used alone, are likely to have shortcomings for enhancing comparability and consistency of interpretation (i.e., safeguarding the standards; Bloxham et al., 2015), over time, and nationally. It appears that social moderation approaches could only be localised in scope, unless they were extended nationally so that the common understandings were shared by all teachers within the jurisdiction who assess those standards (although it is unclear how, or if, this could happen). Even if extended nationally, the collaborative negotiation of meaning may not result in the same interpretation over time, as new assessors enter the community of practice and others exit. For social moderation to enable standards to be effectively safeguarded, it appears to be necessary to combine it with another moderation approach (e.g., external expert moderation, via panel or individual moderator), to ensure comparability. As Hipkins and her colleagues observed, "While social moderation can promote consistency within moderation groups, if there is no external expert to review judgements, it cannot address consistency between moderation groups. This means that it is not really suitable for promoting consistency on a national basis" (Hipkins et al., 2016, p. 124).

In New Zealand, NZQA moderation is conducted primarily using an expert moderation approach, and as already observed (2.4.3, 2.4.4), appears to have a main purpose that relates to accountability. However, beyond the insights that can be gleaned from Ward and Thomas (2016) regarding primary and intermediate schools, there appears to have been no research into the approaches used by organisations in their internal moderation systems.

Thus, few inferences can be drawn from the existing literature about the purposes for which internal moderation is used by different organisations in this country.

2.4.6 Perceptions of other moderation functions

Not all scholars subscribe to the view that accountability and improvement are the main functions of moderation. Two Australian university studies addressed academic staff perceptions of internal moderation. The first (Adie et al., 2013) inductively explored how academic staff thought about and practiced internal moderation.[49] The second study (Grainger et al., 2016) applied in a different context the framework that Adie and her colleagues had developed. These studies found that academics' perceptions of moderation did not align neatly with the prevalent conception of there being two main functions of moderation. Instead, the studies "uncovered the nuance between and within the predominant discourses of learning and accountability" (Adie et al., 2013, p. 972), that illustrated the differing understandings that staff had in terms of the purposes and value of moderation.

In their qualitative exploratory study in one university faculty, Adie et al. (2013) found that staff viewed moderation in four distinct ways. Moderation was seen as promoting 'equity' by ensuring sound and fair assessment that was accurate, consistent, and comparable with standards, across cohorts, between assessors, and over time. Moderation was also seen as 'justification' in that it enables staff to be confident in their own assessment judgements, and to justify their judgements to others (e.g., students) if needed. This perspective of moderation placed the participant as central to the purpose of moderation. The third way that moderation was seen was as 'community building': of collaboration within and across teaching teams to develop shared understandings of standards, the qualities of student work required, and assessment criteria. This view was also associated with staff working together to review and develop assessment tools, learning activities, and teaching approaches to support student learning. Lastly, moderation was seen as 'accountability': as concerned with the distribution of grades for each assessment, and in which the course coordinator was seen to be the expert and final arbiter of grades. This fourth view of moderation sometimes introduced a norming influence, which in turn strengthened assessors' focus on the justification of the grades they had awarded. Moderation as 'accountability' fits neatly into one prevalent conception of moderation (as having accountability purposes), and moderation

---

[49] Beutel et al. (2017, see 2.4.4.1) re-analysed the same data-set as analysed by Adie et al. (2013).

as 'community building' fits neatly into another (moderation as having learning and improvement purposes). However, moderation as 'equity' and 'justification' span accountability, and learning and improvement. Adie and colleagues found that the enactment of moderation within the faculty reflected the dominant view(s) of those involved. Further, none of the 25 participants subscribed to all four view(s) of moderation, and each felt that the view(s) they held, and their moderation practices, were best (Adie et al., 2013).

Grainger et al. (2016), in a small ($n = 10$) qualitative study in a different university, found that casual contract academic staff subscribed to the four ways of viewing moderation, as described by Adie et al. (2013). They found that participants wanted to feel connected to a community of practice, and saw social moderation as providing an avenue through which this could occur. Unsurprisingly then, 'community building' was the view of moderation that was most commonly held. The researchers also found that a lack of continuity in staffing, and the need to build relationships within teaching teams, were perceived as challenges to creating a safe and supportive environment for moderation to occur in, and for developing shared understandings of standards. Grainger and colleagues concluded that these challenges were accentuated for the study participants because the participants were casual contract staff (and as such were employed for limited hours), and there was high staff turn-over, both of which would likely hinder community building.

The studies of Adie et al. (2013), and Grainger et al. (2016), have provided evidence that academic teaching staff within the Australian university sector view internal moderation as having accountability, community building, equity, and justification purposes, instead of having primarily accountability and learning purposes. However, it is unknown whether people involved with moderation in other jurisdictions or sectors also subscribe to these views, or whether they view moderation in different ways. Furthermore, it is unknown whether these views of moderation also apply to external or national moderation. These are further gaps in the literature.

2.4.7 Moderation of internal assessment: Summary

To sum up, moderation is widely mandated and used across education sectors and jurisdictions. However, its implementation is far from uniform, with variation in stage(s) of assessment addressed (see 2.4.0.1), approach taken (expert, social, statistical; 2.4.1), and type of moderation engaged with (internal, between organisations, national; 2.4.2). In addition, there are different purposes for which moderation is utilised, which fall into two

main groups: quality control, accountability, and maintaining public confidence (2.4.3), and organisational and professional learning and improvement (2.4.4). There appears to be widespread agreement among scholars (e.g., Crisp, 2017; Danø & Stensaker, 2007; Gustafsson et al., 2015; Klenowski & Wyatt-Smith, 2014; Newton & Shaw, 2014; O'Connell et al., 2016; Wong & Li, 2010) that while there is sometimes tension between these two groups of purposes, they are not mutually exclusive, and instead, can be complementary (see 2.4.5). However, there is emerging evidence that internal moderation is viewed (at least by some academic staff within the university sector) as having more nuance than simply this duality of purpose, and, instead, is seen in terms of accountability, community building, equity, and justification (2.4.6).

2.5 Enactment of policy within education organisations

Educational policies are set by regulatory authorities or quality assurance bodies to direct, control, and influence the practices and outcomes of organisations (Broadfoot, 2007). Spillane et al. (2002), Braun et al. (2010), and others have recognised that, instead of implementing policies exactly as intended by policy makers, organisations *enact* them in ways and with results that may or may not reflect the policy makers' original intentions. Policy is enacted within and by organisations through a "messy" process of "interpretation" and "translation" (M. Maguire, Braun, & Ball, 2015, p. 486). The resulting shape and form of enactment is influenced by multiple aspects, including the nature of the policy, contextual factors, and the people (the "policy actors") involved on the part of the organisation in the policy work itself (Ball et al., 2011a; Braun et al., 2011, 2010; Hardy, 2015; Hardy & Melville, 2018; M. Maguire et al., 2015; Spillane et al., 2002). As such, "putting policies into practice is a creative, sophisticated, and complex process that is always located in a particular context and place" (Braun et al., 2010, p. 549).

The findings of an ongoing qualitative study into policy enactment in four secondary schools in England were reported by Braun et al. (2010), Ball et al. (2011a), Ball, Maguire, Braun, and Hoskins (2011b), Braun et al. (2011), M. Maguire, Hoskins, Ball, and Braun (2011), and M. Maguire et al. (2015), and used to ground the conceptualisation of policy enactment that those papers presented.[50] The four case-study schools were co-educational comprehensive schools with achievement records mirroring the national average and were situated in

---

[50] This conceptualisation of policy enactment was documented in their book, *How schools do policy* (Ball, Maguire, & Braun, 2012). Sheikh and Bagley (2018) saw Ball and colleagues' conceptualisation of policy enactment as seminal.

diverse locations (inner city, suburbs, small town). Three main data sources from each school were drawn on: textual policy artefacts (e.g., handbooks), semi-structured interviews with various policy actors (including managers, academic leaders, teachers), and observations of within-school policy events, including staff meetings and training (Braun et al., 2010). The study's findings illuminated how policies were enacted within these four organisations, and the roles that textual artefacts, context, and the type of policy played.

The other main work that informs the following discussion is that of Spillane et al. (2002). Spillane and his colleagues drew on a body of empirical and theoretical studies from the fields of cognitive science, social and situated cognition, sociology, and social psychology to develop a cognitive framework for how policy actors interpret and understand— "make sense of"—policy directives (2002, p. 393).

While the sample size of the study reported on by Ball, Braun, Maguire, and colleagues was small (which limits the generalisability of findings), the findings—and theoretical conceptualisation grounded in those findings—suggest policy enactment processes and roles that may occur in other organisations. Further, there is an increasing number of studies from several sectors and jurisdictions (although not New Zealand) that have applied the policy enactment conceptualisation of Ball, Braun, Maguire and colleagues, and the cognitive framework of Spillane and colleagues, and which have validated and built on various aspects of those theoretical conceptualisations.[51]

2.5.1 Enactment involves interpretation and translation

In the initial stages of policy enactment, policies are interpreted within organisations, whereby people attempt to make sense of the policy message, and determine what it means for that organisation, what is required, and how to respond. The interpreted policy message is recontextualised and translated for the organisation (Braun et al., 2010; Spillane et al., 2002). The impact of the interpreted, translated policy message on the organisation is conceptualised and negotiated. For instance, the negotiated impact and response may be to

---

[51] These studies include: The case study by Hardy (2015), which examined the enactment of curriculum reform policy in a large, low socio-economic primary school in Queensland, Australia; the case studies by Dalby and Noyes (2018) into mathematics education policy enactment in three large Further Education Colleges in England; Hardy and Melville's (2018) exploration of how regional senior educators in an education district in Ontario, Canada, engaged with and made sense of a provincial assessment policy; and the grounded theory work by Sheikh and Bagley (2018) that explored the emotional and affective dimensions of teachers engaging in curriculum reform policy work in one Canadian secondary school over a three-year period.

accommodate the policy by modifying existing practices, or to develop new practices to address policy aspects (Braun et al., 2010). Textual artefacts (e.g., organisational policy manuals) are developed to communicate those enactment practices (M. Maguire et al., 2011). The interpretation and translation processes involve creativity and innovation on the part of policy actors who are active in those stages of the enactment process (M. Maguire et al., 2015). Through the enactment process, aspects of the policy may be picked up and attended to, and others may be ignored or not recognised (Spillane et al., 2002). Furthermore, while the official discourses from an organisation communicate how the organisation intends to enact the policy, how it is actually put into practice within the organisation is then interpreted, negotiated, and sometimes contested, by others involved, such as teachers, middle managers, or administrators (Hardy, 2015; M. Maguire et al., 2011). Dalby and Noyes (2018) observed that a single policy was interpreted and translated multiple times by multiple policy actors as it progressed through different levels within a Further Education College (from the senior management team, to faculty level, to department level, and finally into teaching teams), resulting in enacted practices that were inconsistent and divergent.

2.5.2 Impact of status and prescriptiveness of policy

Whether or not a policy is mandated, and how prescriptive it is, impacts on how it is enacted within organisations (Ball et al., 2011b; M. Maguire et al., 2015). Ball, Maguire, and respective colleagues observed that the policies that were "imperative" (Ball et al., 2011b, p. 612) and mandated in nature gave less scope for creativity and imagination on the part of the organisation's policy actors and allowed less opportunity for those actors to contribute to the form of the organisation's policy response. In contrast, policies that were "developmental" or "exhortative" (Ball et al., 2011b, p. 615), or "recommended" (M. Maguire et al., 2015, p. 496), allowed more freedom for an organisation's policy actors to imagine, create, or contribute to the organisational response and form of enactment. For example, in his case study school in which detailed and prescriptive policy directives were being enacted, Hardy (2015, p. 78) observed "a sense of agency-within-constraints" among the staff involved.

In the New Zealand senior secondary and tertiary education sectors, it is a mandated policy requirement that organisations have internal moderation processes in place. However, little is prescribed regarding what form those processes must take, especially for tertiary sector organisations (see 1.2.3, 1.2.4, 2.4.2). It follows that the form of the internal moderation

processes enacted by different organisations across New Zealand is likely to vary. How student samples are selected by organisations for NZQA moderation is another example of the freedom afforded (or not afforded) to organisations to determine how they enact policies by how prescriptive those policies are: TEOs can choose what selection processes they use, where schools are required to use a random selection process (refer 2.4.3.2.1).

2.5.3 Contextual factors mediate policy enactment

The ways in which policies are enacted in organisations are mediated by contextual factors. Braun et al. (2011) found that these mediating factors included organisational context, existing commitments and competing demands, other already-enacted policies, staff expertise and levels, and resourcing and available budgets. The contextual factors were also found to influence how the policy actors interpreted the policy message, and the decisions made regarding the organisation's response (Braun et al., 2011). In his case study school, Hardy (2015) observed how contextual factors including teachers' knowledge of, and assumptions about, the existing knowledge base of their students influenced their responses and decisions regarding how to respond to policy directives. Dalby and Noyes (2018) observed the impact of competing pressures and priorities (in particular, those associated with funding), as well as organisational structures, on policy enactment in the Further Education Colleges they studied. Dalby and Noyes concluded that these contextual factors contributed to the different interpretations and translations of policy message, and subsequently divergent practices, that occurred within the Colleges' various departments.

In the New Zealand situation, where under the current policy imperative of self-management organisations have been encouraged to shape themselves to best meet their communities' and stakeholders' needs (see 1.2.4), it seems probable that each organisation's context will shape how policies are interpreted, translated, and enacted in that organisation.

2.5.4 Policy actors and their roles

The process of interpretation and translation, which leads to enactment, is undertaken by different people—policy actors—associated with organisations. People take on different roles in this policy work (Ball et al., 2011a; M. Maguire et al., 2015). In part, these different roles are linked to the responsibilities held by the person—their "positionality" (M. Maguire et al., 2015, p. 496). For example, an academic leader with responsibility for embedding a

policy response into practice will play a different role in policy work than a classroom teacher with no management responsibilities. M. Maguire et al. (2015) also found that the role people play in policy work is in part determined by their professional knowledge and values and the background they bring. Braun et al. (2011, p. 592) encapsulated this phenomenon with their observation that "[as] policy actors we are always positioned, [and as such] the policy activities we see and how we understand them is dependent on 'where' we are."

Ball et al. (2011a, p. 626) identified different roles that policy actors play in relation to policy work, ranging from a "narrator" who takes the lead in filtering policy, interpreting and explaining it, and communicating the selected course of action to others, a "translator" who develops texts (e.g., handbooks relating to the policy) to assist others with its enactment, to a "receiver" who receives policy directives and adheres to and complies with requirements. The different roles inherently hold different levels of influence or control over the policy enactment: Those who assume narrator roles have more influence over how policies are enacted within organisations than those who assume a receiver role. Furthermore, the policy roles a person takes determines how active that person is at different stages (i.e., interpretation, translation) in the policy work. Those with management responsibility for policy work tend to be the most active in the interpretation and translation stages during which the organisation's 'take' on the policy is being determined (thus they take a narrator role), while teachers, particularly those who are newer or more junior, tend to have a more immediate focus (for instance, on their own classroom and programme), and as such, take more of a receiver role of adherence and compliance with what they are required to do (Ball et al., 2011a; M. Maguire et al., 2015). In his case study school, Hardy (2015) observed different staff assuming narrator, receiver, and translator roles, as they engaged in policy work. Maguire and her colleagues observed of the more senior managers involved in their study, "what comes across is their understanding of the wider context as well as their decision-making capacity—their capacity to interpret and define" (M. Maguire et al., 2015, p. 496), when contrasted with more junior teachers in relation to policy work.

It follows that in the New Zealand education system, policy actors with the positionality to take role of narrators of policies relating to moderation and quality assurance of internal assessment would be those in academic leadership roles within each organisation who have responsibility for and oversight of moderation. Thus, those academic leaders would have the most direct influence over the enactment of the policies relating to moderation and quality assurance of internal moderation within organisations.

2.5.4.1 Cognitive dimensions of policy actors' policy work

Within the policy enactment process, the policy actors who are active in the interpretation and translation stages determine the interpretation and translation that occurs. Spillane and his colleagues posited that policy actors read, decode, and comprehend the policy signal: They "notice, then frame, interpret and construct meaning for policy messages" (Spillane et al., 2002, p. 392). This is an active process. Their 'sense-making' is filtered through and informed by the policy actor's own current knowledge and existing attitudes, beliefs, and understandings; each actor interprets the policy message and constructs meaning based on their existing frame(s) of reference.[52] "What is paramount is not simply *that* [policy actors] choose to respond to policy but also *what* they understand themselves to be responding to" (Spillane et al., 2002, p. 393, italics in original). Thus, different people will notice different things within a policy message and take different meanings from it. Furthermore, people tend to notice and attend to ideas that are similar to what they already know, and as such, may not notice ideas within a policy initiative that are fundamentally different to those which they are familiar, or may (mis)interpret them as being similar. This tendency adds to the likelihood of enactment that diverges from the intention of the policy makers. Spillane and his colleagues posited that the interpretations drawn, and meanings constructed by each policy actor are also influenced by their specific setting, social, historical, and organisational contexts, and by the policy itself. The situated interpretation process heightens the opportunities for the policy intention and message to be transformed by each person in each setting (Spillane et al., 2002).

2.5.4.2 Affective dimensions of policy actors' policy work

Alongside the cognitive dimension of sense-making, M. Maguire et al. (2015) recognised that there is an emotional dimension to people's engagement with policy work, and Sheikh and Bagley (2018) explored this further. Sheikh and Bagley developed a theoretical conception of the emotional, affective components that impact on people's behaviour and response to the policy enactment process. As a core concept, Sheikh and Bagley (2018, p. 55) identified "affective disruption" as the state of emotional imbalance that arises from "the tension created [for a person] by the opposing and competing internal and external forces

---

[52] Hardy and Melville (2018) explored this further and found that not only did the educators in their study make sense of the policy through their existing frames of reference (their knowledge, understandings, experiences, and abilities), but that social-construction within the communities of practice to which the educators had or did belong influenced the development of their frames of reference.

associated with … policy enactment", for example, aspects of a policy that the person perceives as threatening to their own professional autonomy or identity, or questions that the person has over the legitimacy or potential adverse consequences of a policy decision made. This recognition of a highly individualised "emotion-based response mechanism" (Sheikh & Bagley, 2018, p. 58) that acts as a mediating influence on people's engagement in policy work suggests yet another variable that can impact on policy enactment.

2.5.5 Policy enactment within organisations: Summary

The framework developed by Spillane and his colleagues (2002) is instructive because it illuminates the cognitive processes through which policy actors associated with organisations interpret and make sense of policy initiatives, and recognises that this sense-making occurs through the lens of each person's existing frame of reference. The advancement by Sheikh and Bagley (2018) of a construct articulating the affective emotional dimensions to policy work is informative because it recognises the highly personal and individualised emotional responses that people have to policy work, and the influence that these have on their perceptions, understandings, behaviours, and responses. The conceptualisation of policy enactment, as per Ball et al. (2011a, 2011b), Braun et al. (2010, 2011), and M. Maguire et al. (2011, 2015), is influential as it helps to explain how and why the same policy initiative (e.g., a moderation policy) can be enacted in a variety of ways by different organisations, and in ways that diverge (sometimes substantially) from the original policy intent. It does so by demonstrating that policy enactments within and by organisations are shaped in part by the interpretations and translations made by policy actors associated with the organisation, and are situated, contextual, and negotiated. Therefore, the perceptions and understandings of those policy actors who are actively involved in the interpretation and translation of policy directives will influence the policy enactment within and by organisations. In the context of the present study, these policy actors are the academic leaders within each organisation who have responsibility for and oversight of moderation, and the policies of interest are those pertaining to moderation of internal assessment.

2.6 Research gap

Much of the literature holds that the functions of moderation broadly relate to accountability and improvement (2.4.3–2.4.5). However, recent research in the Australian university sector

suggests that internal moderation might sometimes be conceptualised with greater nuance than this (as accountability, community building, equity, and justification; see 2.4.6). While research has been done in the New Zealand secondary context into the perceptions of teachers and principals about the usefulness and value of NZQA moderation (see Controller and Auditor-General, 2012; Hipkins, 2013; Wylie, 2013; Wylie & Bonne, 2016; 2.4.3.5, 2.4.4.1, 2.4.4.2.8.1), none has investigated the perceptions of those in tertiary organisations. No research has been conducted on stakeholders' perceptions in the New Zealand secondary or non-university tertiary sector on the usefulness, value, or functions of internal moderation. Furthermore, no research on stakeholders' perceptions about the functions or purposes of moderation of any type appears to have been done in the New Zealand context, or in the secondary and non-university tertiary sectors.

Educational policy literature holds that policies are enacted by organisations, and that the form and nature of that enactment is determined by how the policy initiatives are interpreted by certain people with positionality to take a lead role in making sense of the policy signal and developing an organisation's 'take' on it (M. Maguire et al., 2015; see 2.5, 2.5.1, 2.5.4). As such, those with the positionality to influence or control each organisation's interpretation and 'take' on policies regarding quality assurance of internal assessment in New Zealand are those academic leaders with responsibility for and oversight of moderation. How the policy actors construct meaning and make sense of policy signals are shaped by those actors' frames of reference and existing knowledge and understanding (Spillane et al., 2002). Therefore, the interpretation of policy directives regarding quality assurance of internal assessment will be shaped by the understandings, perceptions, and existing knowledge of those academic leaders within organisations who have responsibility for moderation.

However, little is known about what academic leaders in different types of education organisations in New Zealand perceive the functions of internal moderation or NZQA moderation to be, or whether academic leaders from different organisation types differ in their perceptions regarding the functions of moderation. This thesis aims to address these research gaps by exploring the perceptions of academic leaders in a sample of ITPs, PTEs, and state schools, and then canvassing the perceptions of academic leaders in these organisation types nation-wide to ascertain any differences. Given the limited amount of research into stakeholder perceptions of the functions of moderation, it was difficult to formulate hypotheses *a priori*, therefore, the research questions are exploratory. In addition, the present study needed to consider the unique context of the New Zealand education sector, and the nuances and differences that exist for the different organisation types within

this. Therefore, it was chosen to contrast three different organisation types[53] that assess against NZQA assessment standards.

2.6.1 Research questions

The research questions that the present study sets out to answer are:

1. What do academic leaders in ITPs, PTEs, and schools, perceive the *in-practice* functions of internal moderation and NZQA moderation to be?
2. To what extent are there differences observable between the three organisation types regarding academic leaders' perceptions of the *in-practice* functions of internal moderation and NZQA moderation?

Therefore, this thesis has taken a pragmatic mixed-methods approach. It has utilised a sequential design incorporating qualitative and quantitative methodologies, conducive to the exploration of perceptions, and then quantification of differences in perception between different groups. The research design used is elaborated on in the following chapter.

---

[53] The scope of the present study was limited to three organisation types to ensure manageability; as a result, other organisation types in the senior secondary and tertiary education sectors were excluded.

3. Methods

To address the first research question (see 2.6.1), it was necessary to ascertain the range of views about the functions of moderation and establish how widely these views were shared across a large sample of academic leaders in education organisations. Answering the second research question required contrasting the perceptions of academic leaders from the three organisation types to identify any statistically significant differences. The same survey instrument and sample as used to answer the first research question were used to answer the second research question. Prior to conducting such a survey, the survey instrument had to be developed, informed by a range of views of the functions of moderation, as held by relevant academic leaders. Thus, a mixed methods exploratory sequential multi-stage research design with parallel data analyses that was based in the pragmatic paradigm was used in this study (Teddlie & Tashakkori, 2009). In the first stage, semi-structured interviews were conducted, and the findings from the interview data analysis were used to inform the development of a survey (the data collection instrument for the subsequent stage; refer Figure 3.1). The second stage of the study involved the collection of quantitative and qualitative data via the survey instrument. Most analyses of the quantitative and qualitative survey data were conducted separately (in parallel), although some integration occurred through these processes: Most analysis of the quantitative data occurred prior to analysis of the qualitative data and informed the deductive analysis of those qualitative data (Creswell, 2012; Teddlie & Tashakkori, 2009).
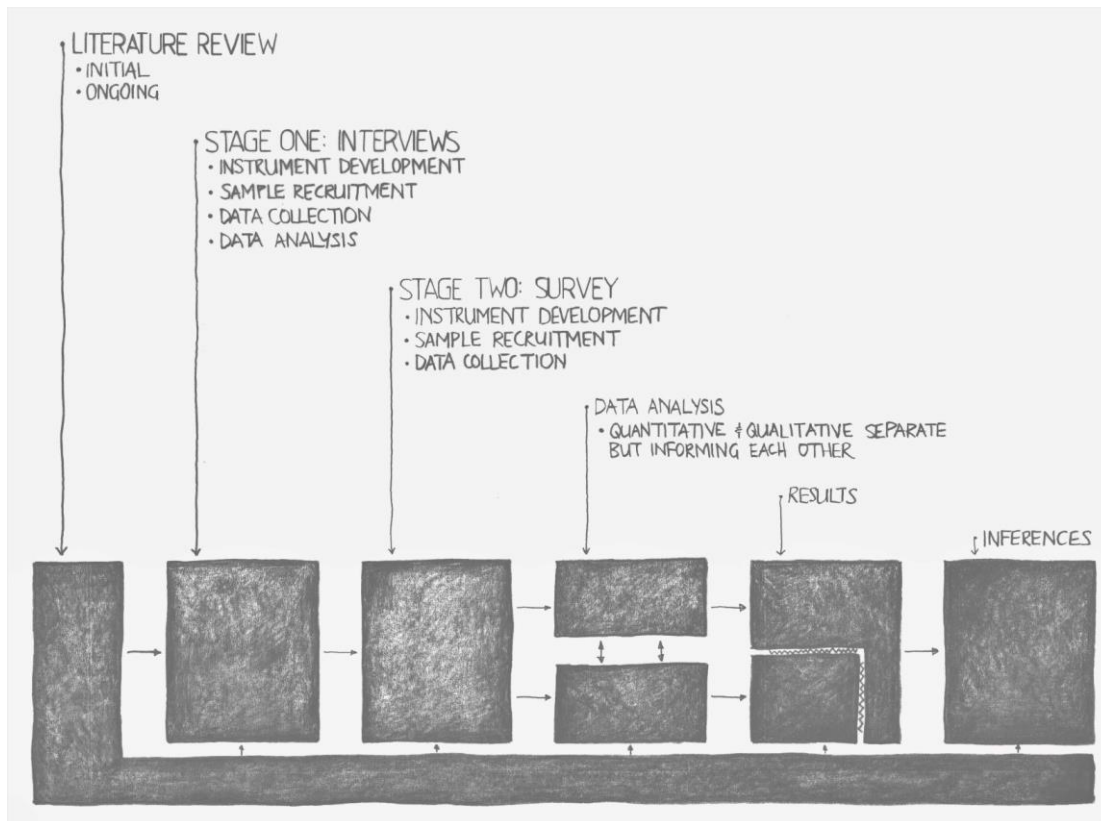
Figure 3.1   Diagram of the present research design

This chapter comprises sections relating to the methods used in each of stages one (interviews) and two (survey), preceded by an explanation of the methodological foundations of the study (3.1), and ethics approval (3.2). Because the purpose of the first stage was to inform the development of the data collection instrument for the subsequent stage, the methods for this stage are initially presented in their entirety (participants and their selection, data collection instrument, procedure, and analysis; 3.3), and a synopsis of the findings are included as Appendix 4. The methods for the second stage are then outlined in section 3.4: Participants (and their selection), then data collection (development then description of instrument, and procedure). Finally, the data analyses used with the quantitative and qualitative survey data are described separately. Limitations of the research design and methods are discussed throughout this chapter, and the main ones are revisited in Chapter 6.

3.1 Methodology

Clarity regarding the philosophies of knowledge that underpin a research project is important, because these philosophies influence what is studied, how the project is conducted, and the interpretation of the findings (Newby, 2010). The present study is ontologically and epistemologically pragmatic.

3.1.0.1 This study is ontologically pragmatic

Ontologically, pragmatism "views knowledge as being both constructed and based on the reality of the world one experiences and lives in" (Teddlie & Tashakkori, 2009, p. 74), and "asserts that there is a single reality and that all individuals have their own unique interpretation of reality" (Mertens, 2010, p. 11). The research questions of the present study (refer 2.6.1) reflected these dual aspects: In exploring participants' perceptions, the first is based on a recognition that each participant had their own interpretation of reality; in observing the extent of differences between the perceptions of participants from different groups of stakeholders, the second recognised the existence of a reality external to the self.

3.1.0.2 This study is epistemologically pragmatic

Epistemologically, pragmatism holds that knowledge "is not about an abstract relationship between the knower and the known; instead, there is an active process of inquiry that creates a continual back-and-forth movement between beliefs and actions" (Morgan, 2014, p. 1049). The research process of the present study contained a continuous and iterative interaction between understanding, action, and outcome on the part of the researcher and her decisions regarding the research process. Any knowledge generated through this research project may be fallible because of the subjective nature of the phenomena of interest (people's perceptions) and because it has been generated from contextually, culturally, socially, and historically situated data that have been analysed and interpreted by the researcher, who brings her own experiences, understandings, and values to the research process (Biddle & Schafft, 2014; Morgan, 2014). Both subjective and objective viewpoints featured at different points in the research process, which is also consistent with a pragmatic epistemological stance (Teddlie & Tashakkori, 2009).

3.1.0.3 Pragmatism guided the research design

Pragmatism takes a practical orientation to problems. It focuses on finding a workable, fit-for-purpose solution in inquiry, where the research questions drive the research design (Biddle & Schafft, 2014; Hammond & Wellington, 2013; Mertens, 2010; Newby, 2010). It allows the use of both inductive and hypothetico-deductive logic, as appropriate to answer the research questions (Teddlie & Tashakkori, 2009). Mixed methods approaches are a natural fit to the pragmatic paradigm: The "need for pragmatism is paramount" and so the influence of philosophy is downplayed (Newby, 2010, p. 47). The research questions of the present study (2.6.1) drove the research design and methods used, resulting in a mixed-methods sequential exploratory design with parallel data analyses, in which qualitative and quantitative tools and approaches were implemented and both deductive and inductive logic utilised.

3.1.0.4 This study takes a pluralistic paradigmatic approach

Pragmatism rejects the either/or dualism of traditional paradigms (e.g., constructivism versus post-positivism), and instead takes a pluralistic approach (Mertens, 2010; Newby, 2010; Teddlie & Tashakkori, 2009). In doing so, it allows "methodological eclecticism" (Biddle & Schafft, 2014, p. 5). As described below, this research project has aspects that drew from constructivism (and did not align with post-positivism), and others that drew from post-positivism (and did not align with constructivism). Taking such an approach enabled the research questions to be addressed.

3.1.0.4.1 It drew from constructivism

An aspect of this study that drew from the constructivist paradigm was its exploratory focus, illustrated by the first research question (see 2.6.1: *"What do academic leaders…perceive the functions…to be?"*), which recognised that academic leaders have their own perceptions and interpretations (Creswell, 2012), and is consistent with there being "multiple, socially-constructed realities" (Mertens, 2010, p. 11)—as is held by constructivism. The use of semi-structured interviews enabled meaning to be co-constructed between researcher and participant, reflecting constructivist epistemology (Teddlie & Tashakkori, 2009). Interview findings informed the development of the survey tool which enabled participants' perspectives to be included in that instrument (Bogdan & Biklen, 2007), and the use of inductive analysis approaches for the interview and qualitative survey data, were also in

keeping with constructivism. Further, it is recognised that the researcher's values will have influenced the selection of survey items, interpretation of results, and inferences drawn (Bogdan & Biklen, 2007; Teddlie & Tashakkori, 2009).

3.1.0.4.2 It also drew from post-positivism

Other aspects of the study draw from post-positivism. These included the second research question (see 2.6.1: "*To what extent are differences observable…?*") which is underpinned by the notion of measurement, a degree of objectivity, and a sense of a reality that is external to the self (Mutch, 2013; Teddlie & Tashakkori, 2009). Notwithstanding the fact that the topic of interest to the present study are people's beliefs (which vary), these notions reflect the epistemological and ontological stances of post-positivism that there is one external reality,[54] which is "knowable within a specified level of probability" (Mertens, 2010, p. 11). The use of an online survey containing multiple closed-response items, and the use of statistical analyses (including inferential statistics; see 3.5.2), are based on approaches and assumptions that are consistent with post-positivism. Also coherent with post-positivism is the aim of collecting sufficient responses through the survey to generate statistically valid and reliable findings with some degree of potential generalisability in the inferences drawn, and the assertion that while the researcher's values are present in the inquiry, the influence of those values in the quantitative aspects of the study could be controlled (Teddlie & Tashakkori, 2009).

3.2 Ethics approval

Ethics approval was sought from the Victoria University of Wellington Standing Committee of the Human Ethics Committee (Reference: 0000020332) and was granted on 18 November 2013 (before any data collection occurred).[55] This approval was for both stages of the research project (interviews and survey). The approved consent form (for the stage one interviews) and information sheets (for stages one and two) are attached as Appendices 1, 2, and 5.

---

[54] In this case, the external reality is taken as sum of individual beliefs in the target populations, which is modelled by the study samples.
[55] The present study was begun as a Master's degree dissertation, before this researcher upgraded her enrolment to a PhD. As such, the documentation associated with the ethics application refers to the initial degree enrolled in.

3.3 Stage one (interview) methods

3.3.1 Stage one participants: Interview sample

3.3.1.1 Selection of interview participants

The organisations from which the interview sample was drawn were ITPs and PTEs (respectively, publicly and privately-owned TEOs that offer technical and vocational education and training; see 1.2.1), and state and state-integrated schools that have senior secondary levels[56] (henceforth called schools). The population from which the interview sample was drawn was the main academic leader in each organisation who held formal responsibility for moderation within the organisation.

Interviews were conducted solely to inform the development of the survey instrument that was to be used in stage two of the study (see 3.4). Interview participants (academic leaders with oversight of and responsibility for moderation in their organisation from ITPs, PTEs and schools) were purposively selected using maximal variation sampling, to gather multiple perspectives (Creswell, 2012). Newby's (2010) advice for clear, documented criteria for the selection of participants was followed to ensured transparency in the sample selection. Characteristics hypothesised to impact on the perceptions of interest for the sample populations were identified (Creswell, 2012; Ingvarson et al., 2005), and organisations that differed on those characteristics were identified and short-listed to ensure a purposive, stratified sample. The characteristics identified for each organisation type were

- ITPs: membership of the Metro Group[57] (one member; one non-member).
- PTEs: one with main business as limited-credit, short-duration training; one with main business as foundation education-type or high-credit, longer-duration training.
- Schools: size of roll (one large; one small); location (one city; one semi-rural or rural); and decile[58] (one high; one low).

The populations of PTEs and schools from which to draw the short-list were also required to be based in the lower half of the North Island (for accessibility purposes: to enable the

---

[56] As they would conduct summative internal assessment against nationally set assessment standards including those quality assured by NZQA, and therefore engage with NZQA moderation (see 1.2.1, 1.2.3).

[57] At the time of data collection, the Metro Group was a consortium of New Zealand's six major metropolitan ITPs, who collaborated on various strategic and operational activities including advocacy and influencing national policy, education programmes, and internal systems and processes (Metro Group, n.d.).

[58] The decile rating of a school is a measure of the socio-economic status of the communities that make up the catchment area of the school, and is used to determine the funding that the school receives (Ministry of Education, 2018c).

researcher to travel to the organisation to conduct interviews face-to-face if that was most appropriate),[59] and all needed to assess against NZQA quality-assured assessment standards.

The population of ITPs from which to draw the short-list were identified from the list published on the NZQA website.[60] Membership of the Metro Group was identified through the Metro Group website.[61]

Possible PTEs to short-list were identified initially via the published list on the NZQA website (filtered using the "search for provider in Wellington region" function).[62] The "TEO in context" sections of published EER reports (also accessed from the NZQA website) were then examined to ascertain the type of business of each PTE in terms of the criteria listed above.[63] Individual PTE websites were also accessed to ascertain their current business in relation to the PTE criteria listed above.

Possible schools to short-list were identified by using the Directory of Schools spreadsheet (current as of 6 November 2013), downloaded from Education Counts website.[64] The spreadsheet was filtered by authority, urban area, education region, decile rating, and total school roll.

The academic leader from each organisation that was invited to participate was the person within the organisation who had oversight of, and responsibility for, moderation. The selection of academic leaders to interview from each short-list was made by convenience sampling: Short-listed organisations were cold-called via telephone, and an invitation to participate in this study was made to the appropriate academic leader. The first two academic leaders from each short-list who agreed to be interviewed was provisionally included in the study.

---

[59] The population from which the shortlist of ITPs was drawn was New Zealand-wide; the small membership of the Metro Group meant that conducting interviews by telephone needed to be an option.

[60] NZQA webpage address identifying all ITPs in New Zealand: http://www.nzqa.govt.nz/providers-partners/about-education-organisations/itps-in-new-zealand/ (accessed on 16 November 2013)

[61] Metro Group website address: http://www.metros.ac.nz/ (accessed on 16 November 2013)

[62] NZQA webpage address using filter to search for providers in Wellington region: http://www.nzqa.govt.nz/providers/results.do?regionCode=110 (accessed on 16 November 2013)

[63] Some of the EER reports were dated, and the information accessed may therefore have also been dated.

[64] Education Counts website – Directory of Schools spreadsheet: http://www.educationcounts.govt.nz/directories/list-of-nz-schools (accessed on 16 November 2013)

3.3.1.2 Interview sample

Six academic leaders, each with oversight of and responsibility for moderation in their organisation, were interviewed: two from ITPs, two from PTEs, and two from state schools that included senior secondary levels. One ITP was a member of the Metro Group and the other was not. Both PTEs were based, or had a branch, in the lower North Island. The main business of one PTE was short-duration, limited-credit training, and the main business of the other was longer-duration, higher credit-value training. Both schools were situated in the lower North Island. One was a high-decile city school with a large roll, and the other was a semi-rural, low-decile school with a small roll.

3.3.2 Stage one instrument: Interview schedule

3.3.2.1 Selection of semi-structured interviews as data collection method for stage one

The purpose of the first stage of the study was to inform the development of a survey instrument, with which to collect data for the subsequent stage. To ensure that the survey instrument would collect appropriate data to address the first research question (see 2.6.1), it was important that the development of the instrument was informed by participant perspectives (Bogdan & Biklen, 2007); that is, as wide a range of perspectives as possible on the purposes of moderation was sought. Semi-structured one-on-one interviews were chosen as the data collection method[65] for the first stage for several reasons. They allowed the researcher to gain "an in-depth understanding of the topic from the [individual] participant's perspective" by engaging in focussed, purposeful dialogue with each participant about the research topic (Mutch, 2013, p. 120). A schedule of pre-established questions was used to guide each interview and provide a level of consistency to the territory covered within the interviews. At the same time, the researcher was able to adapt the questions asked as required, clarify questions or responses as needed, and explore aspects as appropriate to better understand the participant's perspective (Mutch, 2013; Teddlie & Tashakkori, 2009). One-on-one interviews were also comparatively easy to organise, simply

---

[65] Alternative data collection methods considered (with the reasons they were discounted in parentheses) were written questionnaires (do not allow clarification of questions or responses) and focus group interviews (resource-prohibitive and challenging to organise with geographically dispersed participants, also collect data generated through group interaction, where this study's focus was on individuals' perspectives; Lichtman, 2010).

involving finding a time (and, in the case of face-to-face interviews, place) that were mutually convenient to the participant and researcher.

3.3.2.2 Development of interview schedule

A bespoke interview schedule was developed as none suitable was readily available. The development was informed by literature regarding the design and construction of such schedules (including Creswell, 2012; Lichtman, 2010; Mutch, 2013; Teddlie & Tashakkori, 2009). Schedule content was informed by an initial literature review,[66] through which three main functions of moderation had been identified: the quality control of internal assessment (Connolly et al., 2012; Halliday-Wynes & Misko, 2013; S. Johnson, 2013; Vaughan et al., 2012), the maintenance of public and stakeholder confidence in internal assessment and qualification systems (S. Johnson, 2013; Vaughan et al., 2012), and the provision of professional learning opportunities and support for assessors (Colbert et al., 2012; Smaill, 2013).[67] The content was also informed by the researcher's own professional experiences.

The draft interview schedule was trialled in two pilot interviews with participants who were, or had been, academic leaders in an ITP or PTE, (one of whom had also taught in secondary schools). One pilot interview was conducted via telephone, and the other was conducted face-to-face. After each trial, refinements were made to the interview schedule.[68]

Between the data collection interviews themselves, three more minor refinements were made to the interview schedule. These were referring to 'moderation' as 'moderation of assessment' (for clarity), including early in each interview an explanation (by way of an analogy) of what was meant by 'the functions of moderation', and including a question about when NZQA had last conducted an EER (for ITPs and PTEs) or MNA (for schools) on the organisation.

---

[66] That focused on literature from the school and non-university tertiary education sectors

[67] Subsequent review of the literature revealed dominant discourses to hold the main functions of moderation as accountability, quality control, and maintaining public confidence (on one hand), and learning and improvement (on the other). Refer 2.4.3–2.4.4.

[68] It became apparent that participants perceived the *in-principle* functions of moderation to be different from the *in-practice* functions, which precipitated the following changes. To clarify what was being asked, the main questions regarding internal moderation and NZQA moderation were amended to focus on the functions of moderation *in-practice* (e.g., *"What do you think the functions of internal moderation are in your organisation?"*). A follow-up question about each was refined to focus on any other functions *in-principle* that may not currently occur *in-practice* with that type of moderation (e.g., *"In your view, are there any other functions of internal moderation beyond those you have already discussed, and that might not occur in your organisation? If so, what are they?"*).

3.3.2.3 Description of interview schedule

The interview schedule comprised four sections: introduction; internal moderation; NZQA moderation; and the participant's background and their education organisation.[69] The interview schedule is attached as Appendix 3.

The internal moderation section contained two main questions, regarding what the participant thought the functions of internal moderation were in their organisation, and any other functions of internal moderation (that may not occur in their organisation). It also contained prompt questions for use if needed (relating to the three functions of moderation identified in the literature: quality control, maintaining public confidence, providing professional learning opportunities). The NZQA moderation section contained three main questions, regarding what the participant thought the functions of NZQA moderation were, whether they thought that NZQA moderation fulfilled those functions, and any other functions of national moderation (that may or may not be fulfilled by NZQA moderation). It also contained associated prompt questions which covered the same areas as those in the internal moderation section. Reflecting the exploratory nature of the study, the first main question in the two moderation sections was open (*"What do you think the functions of [internal] [national] moderation are […]?"*). These allowed participants to respond however they chose, instead of in a way that had been predetermined by the researcher (B. Johnson & Christensen, 2008).

3.3.3 Stage one data collection procedure: Interviews

3.3.3.1 Data collection (interviews)

The six data collection interviews for the first stage of the study were conducted in a seven-week period through December 2013 and January 2014.

3.3.3.2 Organising the interviews

A mutually convenient time (and place, if face-to-face) for the interview was negotiated between the researcher and participant. Phone interviews were organised for two participants, face-to-face interviews were organised for the rest.

---

[69] Some questions in the last section were tailored for the organisation type (ITP, PTE, school), and were included to assist the researcher to understand each participant's context (Creswell, 2012).

3.3.3.3 Conducting the interviews

All interviews were conducted by the researcher and were audio-recorded with permission from the participant. Face-to-face interviews were conducted in the participant's office. Consent to participate was confirmed before each interview began (via signing of consent forms in the face-to-face interviews and verbal confirmation of consent given[70] in telephone interviews). The opportunity to ask questions about the study was given at the start of each interview.

The interviews were semi-structured and followed the interview schedule. Unscripted questions were asked where appropriate, to probe for understanding, clarify meaning, or gather contextual information to assist with understanding (Creswell, 2012).

3.3.3.4 Transcription of interviews

The researcher personally transcribed each interview using Windows Media Player and Microsoft Word. This allowed the researcher to begin engagement proper with the data at that point (Newby, 2010), and as such, to begin the data analysis process (Mertens, 2010). . So that the interview data could be used to inform the development of a survey tool, full verbatim transcripts were made. This followed Newby's (2010) advice that the type of data needed to fulfil the purpose for which they were gathered ought to guide decisions about how comprehensive transcripts should be.

3.3.3.5 Member checking

Each transcript was emailed to the relevant participant for member checking (Mutch, 2013). Four participants responded confirming that the transcript accurately captured their intent and meaning. Two participants did not respond; the researcher took their lack of response to indicate that those two transcripts were accurate.

---

[70] With scanned copies of the signed consent forms emailed to the researcher post-interview.

3.3.4 Stage one data analysis: Interview data

The data were manually analysed (in Microsoft Word) using deductive and inductive qualitative approaches. Initially, a predetermined overarching coding framework was applied through which to identify participants' perspectives and interpretations within that framework (Newby, 2010). Later, an inductive analysis approach was taken, in which themes were allowed to emerge from the data, unconstrained by the predetermined framework (Newby, 2010). Deductive and inductive phases involved cyclical processes through which codes and categories were developed, applied, refined, and reapplied (Lichtman, 2010; Mertens, 2010; Newby, 2010). Throughout the analysis process, the researcher was aware of the need to continually confront her own bias (Bogdan & Biklen, 2007). Techniques used to do so included researcher awareness of potential bias, the research supervisors reading the transcripts and checking the data analysis, and the researcher repeatedly interrogating the analysis to ensure that the interpretations drawn were accurately founded in the data (Newby, 2010).

The deductive analysis phase comprised two rounds of coding. The overarching coding framework used in this phase had been developed *a priori* (Lichtman, 2010), based on the interview questions, which in turn had been informed by an initial literature review (see 3.3.2.2). The framework consisted of a matrix of the two types of moderation on one axis (internal moderation, NZQA Moderation), and functions of moderation on the other (quality control, maintenance of public confidence, and provision of professional learning opportunities). Thus, six codes—Internal Moderation: Quality Control, NZQA Moderation: Quality Control, and so on—formed the basis of the framework.

Each transcript was coded individually in the first round of deductive coding. The researcher initially re-read the transcript to re-familiarise herself with its content and overall tone (Newby, 2010), before working through the transcript, applying the predetermined coding framework and developing more specific individual codes to describe what was being said. The coding was applied to units of data—individual words or phrases, sentences, or whole paragraphs—using the comment function in Microsoft Word. The codes used and explanatory notes were also recorded in a separate document that served as a codebook (Mertens, 2010). As each subsequent transcript was coded, a constant comparative method of analysis was employed (Lichtman, 2010; Mertens, 2010), whereby the content of the transcript was analysed in light of the codes used previously to ascertain whether there were adequate similarities to use an existing code, or instead to create a new one. In the second deductive round of coding, the initial codes were examined, and those found to cover similar

concepts were clumped together and refined further into categories (Lichtman, 2010; Mertens, 2010; Newby, 2010). Each transcript was then re-coded using the category codes. Throughout the deductive rounds of coding and categorising data, notes were kept recording emergent themes and concepts that did not adhere to the predetermined coding framework (Lichtman, 2010).

The inductive phase of data analysis was undertaken once the deductive phase was completed. Individual transcripts were re-read expressly to identify emergent themes. Codes were developed for the emergent themes, and other transcripts were consulted to explore nuances of each. The codes were refined and applied to each transcript individually and recorded in the codebook. A cyclical process was undertaken, through which codes and categories were refined, reapplied, and so forth in an iterative way (Lichtman, 2010; Mertens, 2010; Newby, 2010).

The need for further refinement to the coding was identified when the data analysis findings were used to inform the content development for the survey tool. These refinements included the elimination of crossover between different category codes, the development of more descriptive and nuanced codes about professional learning opportunities, and refinement of coding about the maintenance of public confidence.

The data selected for analysis, and themes chosen for refinement, were those that were relevant to addressing the research questions and, therefore, to inform the survey development (Cohen, Manion, & Morrison, 2011). Thus, almost no data relating to each participant's context was analysed, as they were not relevant to the development of the survey. Furthermore, the themes not pertaining to the functions of moderation that emerged through the inductive phase were 'parked' and not subsequently incorporated into the survey. Because the sole purpose of this stage of the study was to inform the design of the survey instrument for the following stage, the full results from analysis of the interview data are not presented. Instead, a synopsis of the findings is included as Appendix 4.

3.4 Stage two (survey) methods

Stage two data were collected via an online survey instrument that contained open-field and closed-response items and collected qualitative and quantitative data. Identical items were put to all respondents, apart from those collecting data about respondents' organisations, which were tailored to the organisation types as indicated by respondents at the survey

outset. Those respondents who answered both open-field and closed-response survey items were included in both qualitative and quantitative survey samples, where those who only answered one type were only included in the corresponding sample (see 3.4.1.2).

3.4.1 Stage two participants: Survey sample

3.4.1.1 Selection of populations to survey, and potential respondents to invite to participate

The populations from which the survey sample was drawn were academic leaders with oversight of, and responsibility for, moderation in ITPs, registered PTEs, and state schools that include senior secondary levels. A census approach (Creswell, 2012) was used to invite academic leaders to participate in the survey to maximise the sample size (and therefore, the statistical reliability of the data): An invitation was extended to all ITPs, and all relevant PTEs and schools in New Zealand.

A list of all ITPs and registered PTEs (and a publicly accessible contact name and email for each), was obtained from the Service Support team of the Quality Assurance Division, NZQA. A list of all schools (and the name of the current principal and an administrative contact email address for each) was obtained from the Directory of Schools spreadsheet, current at 1 April 2014 and published on the Education Counts website.[71] From this Directory, Integrated and Not-Integrated State composite (Year 1–15), and secondary (Year 7–15; Year 9–15; Year 11–15) schools were identified.

In total, 18 ITPs, 575 PTEs, and 439 schools were identified as comprising the target populations. An invitation to participate in the study was emailed to the publicly available contact address for each organisation in the target population, attention to the relevant academic leader (see Appendix 8). It was taken on trust that those who participated in the survey were the academic leaders with responsibility for moderation within their organisation.

---

[71] Education Counts website – Directory of Schools spreadsheet: http://www.educationcounts.govt.nz/directories/list-of-nz-schools (accessed on 12 May 2014).

3.4.1.2 Survey sample

The total survey sample comprised 221 academic leaders, of which 215 made up the quantitative sample, and 208 made up the qualitative sample (refer Table 3.1). The quantitative and qualitative samples largely overlapped (with 202 respondents); the data of 13 respondents were included only in the quantitative dataset, and of six respondents only in the qualitative dataset. In each sample, there were almost double the number of PTE respondents as there were school respondents, and only a small number from ITPs.

The quantitative sample comprised those who had completed enough of the survey for their responses to be included in the quantitative analysis (i.e., they completed more closed-response survey items than only question 3; refer Appendix 7), and those who had completed enough of the survey for their responses to be included in the qualitative analysis made up the qualitative sample (i.e., they had entered a response for at least one open-field question). Of the quantitative sample, 176 respondents (82%) had 'finished' and submitted their survey responses, as had 166 (80%) of the qualitative sample. Although the balance (39 respondents for quantitative, 42 for qualitative) had exited the survey instrument before completion, their responses were included in the relevant analyses.

How representative the samples were could not be ascertained because not all respondents provided demographic or organisational data[72, 73] (via Section 4; see 3.4.2.3). Thus, there may be some bias in the data (Creswell, 2012), and as such, caution would be required if generalising present findings to school or PTE populations. However, the sample comprised two thirds of the ITP population, so generalisation to that population would be less problematic.

---

[72] It is noted that sample bias, and no way to determine its extent, is typical of this type of research.
[73] The lack of a complete data set regarding demographic and organisational information also meant that it was not possible to examine whether there were any relationships between these aspects and the responses. Refer 3.3.4 for a description of the quantitative data analyses conducted.

Table 3.1   Survey sample details

	Total			Quantitative			Qualitative		
	n	% total sample	% TP	*n*	% sample	% TP	*n*	% sample	% TP
Total	*221*			*215*			*208*		
ITPs	*13*	6	72	*13*	6	72	*12*	6	67
PTEs	*132*	60	23	*128*	60	22	*123*	59	21
Schools	*76*	34	17	*74*	34	17	*73*	35	17

Note: TP = target population

3.4.1.2.1 Response rates of various organisations

As shown in Table 3.1, the response rate from schools was lower than that of ITPs or PTEs. There are several factors that are likely contributors to this. Firstly, the accessibility of the target population likely had an impact on the response rate from each organisation type. The invitation to participate was emailed to the publicly available contact address for each organisation. Where that email address was not directly to a member of the target population, the present study relied on the invitation being passed on to the appropriate person in the organisation. It appeared that the publicly available contact email for schools (see 3.3.1.1), tended to be an 'administration' address (e.g., the school's main office), as opposed to the principal's nominees (the target population). This may have contributed to the lower response rate for the school population, as compared to the others. Furthermore, the timing of the data collection initially coincided with the end of the school term, which may have also contributed to the lower response rate from schools. The closing date for the survey was extended to attempt to obtain a higher response rate (and reminder emails sent); however, anecdotally, principals' nominees struggled to find the time to participate. Also anecdotally, principals' nominees are asked to participate in research relatively frequently, a factor which may have also decreased the willingness of some to participate.

3.4.2 Stage two instrument: Survey

3.4.2.1 Selection of online survey as data collection method for stage two

To enable the research questions of this study to be answered (refer 2.6.1), an online survey using a cross-sectional approach was selected through which to collect data at one point in time (Cohen et al., 2011; Creswell, 2012). Quantitative and qualitative data could be collected through the inclusion of closed-response and open-field items in the survey instrument (Creswell, 2012). Quantitative data could be analysed statistically to answer both research questions. Taking a cross-sectional approach (in which a common instrument was used, and each respondent completed largely the same survey content) collected standardised quantitative data, allowing direct comparisons to be made between the responses from the three different organisation types (Cohen et al., 2011), as was required to answer the second research question.

An online survey that was designed to be self-administered at participants' convenience, where members of the target population were invited via email to complete the survey, was a time- and resource-efficient approach to administer to a large sample (Cohen et al., 2011). It enabled large-scale data to be collected, allowing inferences to be drawn from the large data set to address the first research question, and statistically valid inferences to be drawn (Cohen et al., 2011) to address the second (2.6.1).

3.4.2.2 Development of the survey instrument

3.4.2.2.1 Development of content

In the absence of a suitable survey instrument being available, a bespoke one was developed. The design and structure of the survey was informed by literature about survey design (including Cohen et al., 2011; Creswell, 2012). The initial draft content of the survey was developed using a deductive approach (Bogdan & Biklen, 2007), before it was refined as informed by the interview findings. A cyclical and iterative process was undertaken to refine the survey content and design (described below) before the survey was constructed in the online platform.

In the initial development, individual survey items were designed to address the functions of moderation (quality control, maintaining public confidence, and providing professional learning opportunities) as informed by literature (e.g., Connolly et al., 2012; Halliday-Wynes & Misko, 2013; Klenowski, 2013; Smaill, 2013), published policy documents from Australia and New Zealand (e.g., Curriculum Council, 2007; NZQA, 2013c; Queensland Studies Authority, 2010), and the researcher's own professional experiences.

The initial draft survey content was refined in light of findings from the interview data analysis (refer Appendix 4), utilising a common mixed methods technique (Newby, 2010), and adhering in part to an inductive approach (Bogdan & Biklen, 2007). Firstly, the initial draft survey items were mapped against the codes that had been developed through the interview data analysis. New survey items were written for the interview codes that were relevant to the research questions but had not already been addressed by the draft items, thereby incorporating interview participants' perspectives into the survey (Bogdan & Biklen, 2007). These new items addressed additional functions of moderation identified in the interviews including: provide an input into self-assessment and self-review;[74] inform the review of student performance; inform teaching; inform an organisation's practice; provide opportunities for learning through moderation feedback, results, or one-to-one support; and give confidence to an organisation's management team or governing body about the teaching and assessment practices that are occurring. Analysis of the interview data had shown that there was a difference between participants' perceptions of the functions currently emphasised in NZQA moderation in practice and what they would prefer was emphasised in NZQA moderation. To reflect this nuance, the item-stems addressing NZQA moderation were changed from rating each function in terms of *its importance*, to rating *the current emphasis* of NZQA moderation on each function.

Draft survey items that neither aligned with the interview data analysis codes nor were well supported by literature were omitted from the survey content. Draft items that were well supported by literature but did not align with any interview data analysis codes were retained and refined (including items suggesting that moderation has a function in building the confidence of assessment designers and assessors; Colbert et al., 2012; Vaughan et al., 2012). Draft survey items that appeared to be repetitious or poorly aligned with the research questions or target population were deleted. Any technical terms were replaced with simpler

---

[74] Self-assessment and self-review are organisational quality assurance processes that schools and TEOs are required to engage in (refer 2.3.1.3.1).

and more direct language, to reduce the degree of variation in respondents' interpretations[75] of the technical terms (Cohen et al., 2011). Consistency in the terminology used was checked,[76] and a downloadable glossary of the terminology used within the survey created (attached as Appendix 6).

3.4.2.2.2 Building of the online instrument

The survey instrument was built using the Qualtrics survey platform. Respondents could withdraw from the survey at any time by closing their internet browser, and the instrument settings applied were such that respondents could move past any item without responding. These features reflect the assertion that a participant has the "right to *withdraw* at any stage or *not to complete* particular items in the questionnaire" (Cohen et al., 2011, p. 377, italics in original). A response reminder (which could be ignored) was applied to one survey item (Q3: *"What type of organisation are you currently an academic leader or manager in?"*), as the usability of the respondent's data relied on this item being answered.

3.4.2.2.3 Pilot of online survey instrument

Usability testing (on personal computer Apple Macintosh and Microsoft Windows operating systems) and piloting of the survey instrument were conducted before it was launched (Cohen et al., 2011). Pilot participants[77] were selected who had been members of the survey target populations in the past, but were no longer so, meaning that they would not be among those invited to complete the survey for data collection (Creswell, 2012). Three who had been in academic leadership positions with responsibility for both internal and national moderation,[78] were recruited as pilot participants. A fourth (who was at the time in an academic advisory position within an ITP but without responsibility for moderation) was also recruited specifically to check the terminology used regarding the external quality assurance regime in place in the tertiary sector (see 2.3.1.3.1).

---

[75] The importance of avoiding the use of technical terms had been illustrated in the interviews where the meaning ascribed to terms including "validity" appeared to differ between participants, perhaps reflecting the ongoing debate about the meaning of this term (refer 2.1.3).
[76] Variation in terminology used had been revealed through the interviews. For example, those in the school sector referred to external moderation conducted by NZQA as 'external moderation', where those in the tertiary sector called it 'national moderation'. In this study, it is referred to as 'NZQA moderation'.
[77] All were known to the researcher.
[78] One each from an ITP, PTE, and state secondary school.

Each pilot participant was emailed a web-link to the online survey, an electronic (Microsoft Word) copy of the survey content, and a feedback form to complete. Feedback was requested regarding: time taken to complete the survey; any web-display issues encountered; appearance, layout, clarity, and wording of questions and response options; and any terminology issues. Feedback received from the pilot resulted in one NZQA moderation item being split into two (becoming *"To provide opportunities for learning from the moderation results in moderation reports"*, and *"To provide opportunities for learning from the feedback in moderation reports"*), and minor refinements being made to the appearance and layout of the content.

3.4.2.3 Description of survey instrument

The resulting survey instrument used to collect data for the second stage of this study was an online, self-administered survey which was scheduled to take up to 30 minutes to complete (see Appendix 7). The survey introduction contained survey information, and links to the Survey information sheet and Glossary of terminology used (refer Appendices 5 and 6). The rest of the survey comprised four sections: 1) Preliminary questions; 2) Internal moderation; 3) NZQA moderation; 4) Questions about you and your education organisation. The survey was designed to be anonymous, unless the respondent chose to include identifiers in their responses, or to enter their contact details at the end of the survey. Where respondents did enter identifying information in their responses, their data was treated as confidential. As is characteristic of self-administered survey instruments (Cohen et al., 2011; Creswell, 2012), it was recognised that not all participants would complete every survey section before closing out of it. To mitigate this, the instrument settings were such that all data were recorded and retained, irrespective of whether respondents 'submitted' their responses.

Each survey section contained a mix of open-field and closed-response items. *Section 1: Preliminary questions* contained one closed-response item (ascertaining which organisation type the respondent was employed in) and one open-field item. The open-field item (question 4: *"Please list the main functions of moderation"*) did not specify the type of moderation to respond about, so unless otherwise specified, the resulting data could only be interpreted as pertaining to moderation-in-general.[79] In not specifying the type of moderation, the item implicitly gave people permission to respond regarding what the functions *in-*

---

[79] As opposed to pertaining specifically to internal moderation or NZQA moderation

principle could be (Hipkins, 2013). This open-field item was situated before any closed-response rating-scale questions, so as "to elicit candid, unrestricted information" (Teddlie & Tashakkori, 2009, p. 235) regarding what functions the respondent thought that moderation-in-general had, in principle.

Section 2: Internal moderation contained four open-field and 10 closed-response items. *Section 3: NZQA moderation* contained four open-field and six closed-response items. Semantic differential rating scale questions were used in both sections to gather ordinal data about respondents' perceptions of the *in-practice* functions of moderation (Hipkins, 2013) in such a way that allowed "a degree of sensitivity and differentiation of response" (Cohen et al., 2011, p. 386). For example, a stem question of "*Please rate each of the following statements in terms of its importance as a function of internal moderation*" was used in Section 2, with a four-point rating scale of *High importance, Medium importance, Low importance, No importance*. Each closed-response item in these two sections comprised between three and six semantic differential rating scale sub-items that shared the main stem. The open-field items in each section were positioned after each block of closed-response questions and provided respondents with the opportunity to add any functions that they felt had been omitted by the closed-response items. Thus, the open-field items mitigated against the risk that the functions addressed by the closed-response items were not exhaustive (Cohen et al., 2011). These open-field items pertained to what the *in-practice* functions of internal moderation or NZQA moderation were (Hipkins, 2013); respondents were constrained by the positioning of the items in reference to the actual type of moderation.

Eight items that were asked of all respondents comprised *Section 4: Questions about you and your education organisation*. This section also included between three and six items that were tailored for each of the three organisation types and displayed based on the organisation type selected at the start of the survey. The items in this section were a mix of open-field and closed-response (dichotomous and multi-choice) questions.

3.4.3 Stage two data collection procedure: Survey administration

3.4.3.1 Data collection (survey)

The online survey was activated for a period of nine weeks: Wednesday 14 May–Friday 18 July 2014.

3.4.3.2 Distribution of invitation to participate

An invitation to participate in this study and complete the survey was addressed to the academic leader with oversight of and responsibility for moderation, and emailed to all ITPs, PTEs, and schools that were identified as belonging to the target populations. The email invitation (tailored for each organisation type; see Appendix 8 for an example) contained a link to the online survey, to keep the survey responses anonymous and to avoid any download- or layout-problems that are often associated with emailed questionnaires (Cohen et al., 2011). Attached to the email invitation was the *Survey information sheet*, as approved by the Ethics process (Appendix 5).

3.4.3.3 Ongoing maintenance of survey

Throughout the period that the survey was activated, the target population email address list was maintained. The names of those who had included contact details in their survey response, and those who had notified the researcher that they had completed the survey or did not wish to participate, were tagged as such so that no follow-up reminders would be emailed to them.

3.4.3.4 Reminder invitation

Two and a half weeks after the initial invitation to participate was emailed to organisations, a follow-up reminder invitation (tailored for each organisation type; see Appendix 9 for example) was sent.

3.4.3.5 Extension of survey closing date: Second reminder invitation

The survey was kept open for three weeks beyond the original schedule (from the last week of Term 2 until the end of the school holidays), to try to increase the school response rate. Four days after the initially scheduled closing date, a second follow-up reminder invitation was emailed to the target population (tailored for each organisation type; refer Appendix 10 for example), informing them of the extension of survey closing date, and again inviting them to participate.

3.4.4 Stage two data analysis: Quantitative survey data

3.4.4.1 Overview of quantitative data analysis conducted

The quantitative survey data collected (via the closed-response items) were predominantly nominal and ordinal. To contribute to answering the first research question (see 2.6.1), exploratory factor analyses and principal component analyses were run (Tabachnick & Fidell, 2013), and item response distributions were examined (Cohen et al., 2011). To answer the second research question (see 2.6.1), the data for the set of items associated with each factor were calibrated onto (interval) measurement scales using item response analysis (Hambleton, Swaminathan, & Rogers, 1991). One-way analysis of variance tests and independent-sample t-tests were then run comparing organisation types on these measurement scales. Chi-square tests (Cohen et al., 2011) were run on survey items that did not associate with any factors, and which were therefore not calibrated onto measurement scales. Pearson's correlation coefficients were calculated between the measurement scales, and two-proportion z-tests were run on matched pairs of survey items, to assist in addressing the first research question.

Most quantitative data analyses were completed using the software programme SPSS. Calibration of items onto measurement scales was completed using the software programme R. Chi-square tests and z-tests and were completed using Microsoft Excel.

3.4.4.2 Principal component analyses

Principal components analyses with variamax rotation were used as a preliminary technique before exploratory factor analyses to reduce the variables into a smaller number of uncorrelated uni-dimensional components (Bryant, 2000; Manly, 1986). These principal components analyses were conducted to indicate the likely number and nature of factors that existed within the data and would be extracted through the subsequent exploratory factor analyses[80] (Tabachnick & Fidell, 2013). Principal components analyses involving data from the items pertaining to internal moderation were run separately to those involving data from the NZQA moderation items. The results of these principal components analyses have

---

[80] Principal component analysis was only appropriate to use as a preliminary technique. It would have been an inappropriate technique to use to answer the first research question (which assumes the presence of underpinning latent traits), as no theoretical analysis is conducted through it (Tabachnick & Fidell, 2013).

not been reported in Chapter 4 (Results), as they were preliminary to and superseded by exploratory factor analyses (3.4.4.3).

3.4.4.3 Exploratory factor analyses

To contribute to answering the first research question (2.6.1), exploratory factor analyses were conducted on the quantitative data. These analyses enabled the latent structure underlying the variables to be ascertained and understood, and underlying latent traits or constructs to be theorised (Bryant & Yarnold, 1995; Tabachnick & Fidell, 2013). Exploratory factor analysis was an appropriate statistical technique to use because the first research question was exploratory and akin to the development of a hypothesis or theory;[81] it was assumed that there were underlying constructs that produced the observed scores on the variables and this study aimed to identify them (Tabachnick & Fidell, 2013). Conducting exploratory factor analyses was also a preliminary step to the calibration of items onto measurement variables, which was to enable the second research question (2.6.1) to be addressed.

The maximum likelihood extraction method[82] and the oblique rotation procedure[83] Oblimin with Kaiser Normalisation were used in the exploratory factor analyses, and initial analyses were set to retain eigenvalues of greater than one. Separate exploratory factor analyses were conducted on data from the internal moderation survey items and from the NZQA moderation survey items.

When evaluating the analysis solutions, the researcher took a parsimonious approach, aiming to retain only sufficient factors to account for as much variance in the data as possible (Tabachnick & Fidell, 2013). Decisions about which items and factors to retain, and whether or not to specify the number of factors to be extracted, were informed by the scree plot, pattern and structure matrices, and how interpretable the data were (Tabachnick & Fidell, 2013). Factors above the 'break' in the scree plot of eigenvalues were retained. The

---

[81] In contrast, confirmatory factor analysis is used to *assess* theories or *test* hypotheses, and as such, would not have been an appropriate technique for the first research question.

[82] As appropriate for an exploratory study, because it increases the probability that the factor loadings estimated from the sample data will best represent the characteristics of the population (Tabachnick & Fidell, 2013).

[83] Such procedures allow the factors to correlate or not to the extent determined by the data (unlike orthogonal procedures which force the factors to be uncorrelated; Tabachnick & Fidell, 2013). It was assumed that the factors involved were likely to be correlated to some extent, given that they pertained to people's perceptions.

pattern matrix showed more than one item contributing primarily to each retained factor (and at least one item loading strongly onto each). Furthermore, each retained factor needed to "make sense" (Tabachnick & Fidell, 2013, p. 613): The items shown in the pattern matrix as contributing to the factor shared a conceptual link, which was distinct from any other factor in the solution. The final consideration used to determine whether to specify the number of factors to extract in an analysis was whether each factor could subsequently be shown to comprise only one dimension. This was determined by conducting separate principal component analyses on the items associated with each factor.

In the exploratory factor analyses run on the survey items pertaining to internal moderation, one item did not show as loading onto any factors in the pattern matrix generated from the initial analysis and so was omitted from subsequent analyses. When the considerations outlined above were applied, nine factors were retained.

In the exploratory factor analyses run on the NZQA moderation items, one item did not show as loading onto any factors in the pattern matrix generated from the initial analysis, and a second item was conceptually incongruent with all other NZQA moderation items. Both items were therefore omitted from subsequent analyses. When the considerations outlined above were applied, five factors were retained.

3.4.4.4 Principal component analyses on each factor

The dimensionality of the factors identified through exploratory factor analyses were checked prior to calibration onto measurement scales by running a separate principal components analysis with varimax rotation on the items associated with each factor. Each principal component analysis was set to retain eigenvalues greater than one. For each internal moderation factor, only one component was extracted, showing that the factor was uni-dimensional. For four of the NZQA moderation factors, only one component was extracted from each factor, showing them to be uni-dimensional.

For the other NZQA moderation factor, two components were extracted, although the principal component analysis showed the factor to be *almost* uni-dimensional. The eigenvalue of the second component (1.063) was only just above the cut-off point of one, and only two items showed as loading against this component in the rotated component matrix. In the unrotated component matrix, those two items each loaded strongly onto both components. The unrotated component matrix showed one item as loading more strongly

onto the first component (with a loading of 0.712, compared to a loading of 0.558 onto the second), but showed the second item as loading slightly more strongly onto the second component than onto the first (with a loading of 0.659 compared to 0.631). It was considered that this factor was close enough to being uni-dimensional to proceed with calibrating the items onto a measurement scale.

3.4.4.5 Items calibrated onto measurement scales

The data for items contributing to each factor were calibrated onto a measurement scale using a one-parameter graded response model, an item response theory model (Samejima, 1969; see also DeMars, 2010; Hambleton et al., 1991). Model parameters were estimated using a maximum likelihood iterative algorithm. The calibration onto measurement scales was completed to allow parametric tests including one-way analysis of variance and independent sample *t*-tests to be validly conducted, to answer the second research question of this study (see 2.6.1).

Integral to item response theory is the premise that a person's response to (or performance on) an item can be explained by a set of traits, abilities, or attitudes (in the case of this study, their beliefs about a construct, e.g. the importance of a certain function of internal moderation). Also integral is the premise that the relationship between a person's response to an item and their attitude underlying that response is such that as the level of their attitude increases, the likelihood of them having a certain response also increases. This relationship can be described by a "monotonically increasing function" (Hambleton et al., 1991, p. 7).

Item response theory is underpinned by the assumption that there is only one dimension to the construct that the measurement scale is against (DeMars, 2010; Hambleton et al., 1991). This assumption was met: Each factor was shown to be uni-dimensional (or close enough to be considered uni-dimensional) through principal component analyses (see 3.3.4.4).

3.4.4.6 Mean by organisation type on measurement scales

The quantitative data were disaggregated into organisation type (ITPs, PTEs, schools) according to how each respondent answered question 3 (see Appendix 7). In preparation for answering the second research question (2.6.1), the mean and standard error of the mean for each of the three different organisation types were calculated for each measurement

scale. The locations of the means and standard errors were then plotted onto histograms, to allow visual inspection of their locations and any differences between them.

3.4.4.7 One-way analysis of variance tests and independent sample t-tests

A one-way analysis of variance test was conducted on each measurement scale to ascertain whether there were statistically significant differences in the mean locations of the three organisation types (Tabachnick & Fidell, 2013). Where a statistically significant difference was found, independent sample *t*-tests were conducted on the three possible pairings of organisation types, to ascertain which organisation types the statistically significant differences were between (Manly, 1986). These tests enabled the study's second research question (2.6.1) to be answered. It was noted that the *t*-tests involving ITPs were of low statistical power, due to the small ITP sample size compared to the sample size of the other two groups (Manly, 1986).

3.4.4.8 Chi-square tests for independence

To contribute to addressing the second research question (2.6.1), chi-square tests for independence were run on those items omitted from the exploratory factor analyses (and therefore not calibrated onto measurement scales; see 3.3.4.3) or that contributed to a factor where the items subsequently would not calibrate onto measurement scales. The chi-square tests compared response distributions for each item by organisation type; as data contained in these items were ordinal, chi-square tests for independence were valid statistical tests to run (Cohen et al., 2011).

The null hypothesis tested in the chi-square tests was that there was no relationship between the response categories selected and the organisation type or moderation type respectively. Critical χ^2 values for the various degrees of freedom at the .05 and .01 levels were accessed from Larsen and Marx's *Table A.3 Upper and Lower Percentiles of χ^2 Distributions* (Larsen & Marx, 1986, p. 581).

The chi-square tests for independence run on individual items compared the response distributions for PTEs to those of schools. ITPs responses were not included in the tests due to the low ITP sample size. In tests showing expected frequencies of less than five for any

response category for either group (e.g. by PTE respondents for the response category of *No importance*), that response category was combined with the neighbouring category (in this example, *Low importance*) and another chi-square test was conducted (McDonald, 2014). Where there was still an expected frequency of less than five for any response category for either group in the subsequent test, the test was abandoned because it would be statistically unreliable (McDonald, 2014). Any tests with only one degree of freedom were also abandoned because they indicated that the ordinal data were clustered across two neighbouring response categories, and, therefore, little in the way of substantive interpretation could be made.

3.4.4.9 Pearson's correlation coefficients between measurement scales

Pearson's correlation coefficients (2-tailed) between measurement scales were calculated to ascertain the nature of any correlations between the measurement scales (Cohen et al., 2011). This was to contribute to addressing the first research question (see 2.6.1). Correlations between internal moderation scales, between NZQA moderation scales, and between internal moderation scales and NZQA moderation scales were calculated. Only those that were statistically significant at the .05 level or above were retained.

Pearson's correlation coefficients (2-tailed) were also calculated between measurement scales and the number of years that respondents had been involved in education (in academic management, and as a teacher, assessor, or administrator), to ascertain the nature of any correlation between these variables (Cohen et al., 2011). Only those correlations that were statistically significant at the .05 level or above were retained.

3.4.4.10 z-tests for proportions

In an example of the "iterative cycle between inductive and deductive approaches" that characterises mixed methods research (Teddlie & Tashakkori, 2009, p. 80), a finding from the analysis of qualitative survey data prompted further quantitative analysis of certain survey items: *z*-tests for proportions were run on eight pairs of items that asked about identical functions of internal moderation and NZQA moderation.[84] The *z*-tests compared the

---

[84] For example, item 6.1, *Internal moderation: To check that assessments are fair to all students,* and item 21.4, *NZQA moderation: To check that assessments are fair to all students.*

perceived importance of a function of moderation conducted internally with the perceived emphasis placed on the same function by NZQA moderation.

Organisations do not have control over NZQA moderation or the functions that it emphasises. However, those in organisations who are in policy narrator roles—such as the academic leaders with responsibility for moderation (i.e., this study sample)—are likely to have influence over internal moderation in their organisation, including the functions that are emphasised through it (Braun et al., 2010; M. Maguire et al., 2015; Spillane et al., 2002; see 2.5.4). Drawing on the policy enactment research traversed in section 2.5, the assumption was made that the importance of functions of internal moderation, as perceived by academic leaders in the sample, would likely be reflected in the emphasis placed on those functions through internal moderation in their organisations. Based on this assumption, the z-tests were conducted.

For each pair of items, the proportion of respondents rating internal moderation as being of *high* importance was compared with the proportion rating NZQA moderation as having *high* emphasis, and the proportion rating internal moderation as being of *medium* or *high* importance was compared with the proportion rating NZQA moderation as having *medium* or *high* emphasis. The null hypothesis was that the proportions rating the items within each pair were equal, with the alternative hypothesis being that the proportions were not equal (Sirkin, 2006). The *p*-values involved were two-tailed. The z-tests contributed towards addressing the first research question (2.6.1).

3.4.4.11 Distribution of response frequencies for each item

The distributions of response frequencies across the categories for each item were also considered and contributed to addressing the first research question (2.6.1). For each item, the modal information was examined (Cohen et al., 2011).

3.4.5 Stage two data analysis: Qualitative survey data

3.4.5.1 Overview of qualitative data analysis

The qualitative survey data were analysed using both inductive and deductive approaches, to assist addressing the first research question (2.6.1). It was initially planned that the

qualitative data from the internal and NZQA moderation survey sections would be deductively analysed using a predetermined coding framework based on findings from the quantitative survey data. However, initial application of that predetermined framework showed that some of those codes did not 'fit' the qualitative data well, and instead a more inductive approach (in which codes were developed to 'better fit' the data) was required. The planned approach was modified,[85] and data from the internal and NZQA moderation sections were analysed using both deductive and inductive approaches. Question 4 of the survey had generated data about what respondents perceived to be the *in-principle* functions of moderation-in-general, and these were analysed using an inductive approach.[86] Data were coded without reference to the organisation type of respondents. Only once all coding was complete were the coded data disaggregated into organisation types, to assist in addressing the second research question (see 2.6.1).

In keeping with the characteristics of qualitative data analysis, the analysis process was iterative and cyclical (Mertens, 2010; Newby, 2010). Most of the analysis was conducted one survey section at a time (first internal moderation, then NZQA moderation, then moderation-in-general); however, the analysis of each section informed refinement of the analysis of the other sections in an iterative way. Further refinement of the coding and coding framework occurred as the analyses of the three survey sections were brought together. (Refer Appendices 11-13 for the final versions of the coding frameworks.)

Responses to open-field items in online surveys are often characterised by their brevity, or by being written in 'note form' or incomplete sentences. As such, the meaning of these responses can be ambiguous (Cohen et al., 2011). Furthermore, it is not generally feasible to check meaning or to seek explanation regarding responses from survey respondents (Cohen et al., 2011; Mutch, 2013). Thus, the researcher must analyse collected data without benefit of additional clarification. Several tools and approaches were utilised to provide transparency in the interpretations made through the present qualitative analysis process and to assist coding consistency. These included the use of coding guidelines, inclusions lists, and coding sheets, as well as repeated interrogation of the coding framework, guidelines, and lists, and taking a systematic approach throughout. These tools and approaches are now described.

---

[85] The modification of approach reflects Newby's (2010) and Teddlie and Tashakkori's (2009) observations that mixed methods research necessitates having sufficient flexibility in the research design to adapt the design in response to what arises during the study.
[86] Exploration of academic leaders' perceptions of the *in-principle* functions of moderation became a *post hoc* research question for this study. Refer 4.4 and 5.0.

Coding guidelines and inclusions lists were developed and utilised to help ensure consistency in the coding and transparency of interpretation. (Refer Appendices 14 and 15 for extracts.) The coding guidelines[87] were developed for the analysis of the internal moderation data during initial coding of that data (refer 3.4.5.3) to inform decisions made during the coding process. The guidelines were then adopted for the analysis of the NZQA moderation data (see 3.4.5.4), before being adapted for the analysis of the moderation-in-general data (see 3.4.5.5). An inclusions list was developed for each code to define what was included in (and, for some, what was excluded from) the code, and assist the differentiation between codes when applying them. The quantitative survey items associated with each factor (identified through quantitative analysis, refer 3.4.4.3) formed the basis of the inclusions list for each predetermined code. The inclusions lists were expanded as appropriate additions were identified through engagement with the data. For the other codes, inclusions lists were developed and refined as new codes were introduced and applied. All inclusions lists were treated as 'live' throughout the coding process: Additions and amendments were made to the relevant lists as decisions were made about how data were to be coded. Each time a change was made to an existing code, the relevant inclusions list was reviewed and amended if necessary. Amendments to inclusions lists for codes for one section of data (e.g., moderation-in-general) also precipitated amendments to the inclusions list for a related code from a different section (e.g., internal moderation).

Through the data analysis process, the coding frameworks, codes, and inclusions were interrogated numerous times and refined where necessary to ensure that the analysis findings credibly reflected the data.[88] This was achieved by the researcher keeping notes throughout the analysis process of observations made, data that 'didn't fit' with the existing coding framework, and conceptual misalignments within the codes. These notes, alongside the ongoing record kept in the coding sheets (refer below), informed the ongoing interrogation and refinement.

In accordance with Lichtman's (2010) and Newby's (2010) advice, the data were worked with in a systematic way. For example, each time a refinement or amendment was made to a code, inclusions list, or coding guidelines, another sweep was made through the data or relevant coding sheet to review the existing coding in light of the change, or to apply the next iteration of code or inclusions. Multiple readings of the survey responses throughout the analysis process ensured close engagement with the data (Mertens, 2010; Newby, 2010).

---

[87] These comprised instructions such as, "where a response contains different aspects that related to different codes, each aspect will be coded to the relevant code".

[88] As opposed to being what the researcher wanted to find (Newby, 2010).

The qualitative analysis was conducted using NVivo 11 for Windows. Coding guidelines, coding frameworks, and inclusions lists were stored in Excel spreadsheets (which functioned as a code book; Mertens, 2010). The spreadsheets also contained coding sheets to assist with the "house-keeping" (Newby, 2010, p. 465) of the coding process. The coding sheets comprised matrices into which the coding of individual respondents' data was recorded as it occurred, to assist in the monitoring and review of the coding process. Once themes that were shared across the three survey sections were identified (refer 3.4.5.6), close analysis of the associated data was conducted manually, and hard copy records of this detailed coding maintained. The coding sheets, hard copy records, and iterations of coding guidelines and inclusions lists, provided an audit trail of the decision-making that occurred throughout the qualitative data analysis and "[made] visible the decision-making trail" (Mertens, 2010, p. 429).

3.4.5.2 Description of the qualitative data analysis process

The qualitative data analysis process used is described below, to illustrate the care taken, and provide confidence in the inferences drawn (Mutch, 2013). The actual process contained more iterations than are reported here. For the sake of brevity, not all cycles of the process have been described below, although they are all documented in the code book.

Initially, all qualitative data were cursorily read, to enable the researcher to get a sense of the content and scope of the data (Newby, 2010) and to determine the suitability of the planned analysis approach.

3.4.5.3 Internal moderation section

The first set of qualitative data analysed was that from the internal moderation section of the survey (see 3.4.2.3; Appendix 7). A deductive approach was employed in the first instance, whereby the nine factors identified through exploratory factor analysis of the quantitative data (refer 3.4.4.3) were used as the initial coding framework.

The internal moderation data were read with the initial codes in mind, before the first tentative coding was undertaken. During this first round of coding, it became apparent that some of the initial factor-based codes were more specific and narrower than the data, rendering those codes inapplicable without making unwarranted assumptions about

respondents' meaning.[89] To address this type of issue, a semi-inductive approach was adopted, whereby broader codes were developed to complement the initial factor-based codes (e.g., the broader code *Internal Moderation: General assessment quality* was established to complement the more specific codes). Global-level category codes were instituted to encompass all codes that addressed a certain broad topic (e.g., *Assessment quality*, which encompassed the broader code and three specific codes cited in the immediately preceding footnote, plus others that were later developed). It was also found that two of the factor-based codes (*Internal moderation: Assisting organisational development* and *Internal moderation: Providing professional learning opportunities*) did not align closely enough with much of the data to ascertain which code was most appropriate (e.g., *"Collegial conversations and learning do not happen in this rushed and overworked environment"*). To address this issue, the two codes were collapsed into one: *Internal moderation: Organisational and professional learning and development.* See Appendix 11.

3.4.5.4 NZQA moderation section

The analysis process used for the dataset generated from the NZQA moderation section mirrored the process undertaken with the internal moderation data (3.4.5.3): An initial deductive approach using a predetermined coding framework based on the quantitative findings (3.4.4.3) was then modified to a semi-inductive approach. The lack of specificity in some NZQA moderation data precipitated the incorporation of additional codes into the coding framework that were broader than the predetermined codes (e.g., *Assessment quality*). Another predetermined code was broadened[90] to better represent data relating to learning opportunities (e.g., *"NZQA moderation doesn't necessarily lead to professional learning opportunities in my opinion…"*). A fully inductive approach was also taken, resulting in the identification of several emergent themes. Further amendments were made to the coding framework with the addition of codes to encapsulate them, including *Educational quality* and *Dissatisfaction with NZQA or regulatory context*. See Appendix 12.

---

[89] For example, it was not possible to ascertain which, if any, of the following codes the response, *"Internal moderation is a vital part of the structure that ensures efficient and robust facilitation of assessment"* was most closely related to: *Checking assessment material quality*, *Checking assessor judgement quality*, or *Improving assessment material and assessor judgement quality.*
[90] From *Providing opportunities to learn from NZQA moderation reports* to *Learning opportunities provided by NZQA moderation*

3.4.5.5 Moderation-in-general section

An inductive approach using a constant comparative method (Teddlie & Tashakkori, 2009) was taken to the analysis of the moderation-in-general data (see 3.4.2.3). The analysis was conducted after most analyses of the internal and NZQA moderation data had been completed, although aspects of the moderation-in-general analysis were used to inform further refinements in the previous coding.

The coding framework for the moderation-in-general data was developed through a process of reading the data and recording in diagrammatic form the researcher's impressions of the meaning those data contained. These diagrammatic representations were refined through a number of iterations,[91] each in light of further readings of the data, until key topics were revealed (Lichtman, 2010; Mertens, 2010) and codes assigned. Codes from the frameworks employed for the internal and NZQA moderation sections data were incorporated into the framework where appropriate (e.g., *Educational quality*). To resolve conceptual difficulties, the existing moderation-in-general coding framework was set aside at a later juncture in the analysis process, the data were read afresh, and notes were taken of key emergent themes. This process confirmed much of the existing framework, but also resulted in the development of codes relating to the assessment properties with which moderation is concerned (see Appendix 13).

The moderation-in-general coding framework included a structure for the general topic of *Assessment quality* that provided three dimensions against which to code those data: The **assessment component** that moderation acts on (e.g., assessment materials or assessor judgements), the **action that moderation takes** on the quality of assessment components (e.g., assuring, or evaluating it),[92] and the **assessment properties** with which moderation is concerned (e.g., fairness, consistency).

For transparency (Mutch, 2013), further explanation regarding the development of the inclusions lists for the three dimensions of *Assessment quality* codes is provided here. The lists for the **assessment component** codes (e.g., *Assessment materials*) were based on the various terms used by respondents for each component. For example, responses that referred explicitly to assessment "materials", "tasks", "tools", "instructions" (or similar) were

---

[91] Refer Appendix 16 for examples of iterations.
[92] Instituting codes addressing the actions that moderation takes was in preparation for triangulation of the qualitative and quantitative findings (Newby, 2010; Teddlie & Tashakkori, 2009): The previously conducted quantitative data analysis (3.4.4.3) had found that survey items addressing *checking* assessment quality were distinct from those addressing the *improvement* of assessment quality.

included in the *Assessment materials* code. The inclusions lists for the **action that moderation takes** codes (e.g., *Assure assessment materials*) were constructed based on the verbs that respondents had used to describe the functions of moderation in relation to assessment quality. Categories of verbs that appeared in the data (e.g., "ensure") were compiled, using their definitions as given by the Concise Oxford Dictionary. For example, the verb "ensure" was defined as "make (person, thing) safe, (*against* risks); make certain (thing, *that*); secure (thing *to, for,* person etc)" (Sykes, 1982, p. 321, italics in original), and it was therefore categorised with the other "assure" verbs, and incorporated into the inclusions list for the *Assure…* moderation action codes. The inclusions list for the **assessment properties** codes with were based on the terms used or meaning contained in the data. Specification of what was excluded from each code was important for differentiation between the assessment property codes. For example, "ascertaining consistency with the national standard" was included in the code *Assessment assesses what it is meant to assess* and not the *Assessment is consistent* code. The properties of *validity* and *reliability* were allocated their own stand-alone codes,[93] and data were coded to them when respondents explicitly identified "validity" or "reliability" as properties of assessment that moderation is concerned with, to ensure that respondents' intended meanings were not misinterpreted through the data analysis process.

3.4.5.6 Bringing the qualitative analysis of the different survey sections together

Once the coding of qualitative data was completed, themes of associated codes occurring across moderation-in-general, internal moderation, NZQA moderation were identified (Lichtman, 2010; Newby, 2010), for example, *Assessment quality*. For each theme, close analysis of data from the associated codes was undertaken using an inductive approach of detailed coding to ascertain what the data were 'saying'. Only once the close analyses had been conducted were the coded data disaggregated into organisation type, enabling comparisons to be made between the organisation types for the qualitative data, as was required to answer this study's second research question (see 2.6.1).

---

[93] The meaning of the term 'validity' is contested, and the term 'reliability' has a specific technical definition (refer 2.1.3). It was unclear what meaning respondents had ascribed to the term 'validity' when they included it in their answers, or whether their intended meaning of 'reliability' cohered with its technical definition.

Care was required to ensure that inferences drawn from the findings from the different survey sections were warranted, because of the nature of the open-field item(s) in each section (refer Appendix 7). The moderation-in-general item (question 4; see 3.4.2.3) was asked at the start of the survey and was designed to elicit top-of-mind responses. Thus, inferences could be drawn from the number of respondents identifying each function about the level of awareness of that function, and its urgency and presence for respondents (in terms of what 'sprang to mind' when they thought of moderation). In contrast, the open-field items in the internal moderation and NZQA moderation sections were complementary to the closed-response items and were framed as opportunities to add anything further to those aspects covered by the closed-response items (e.g., question 10: *"Please describe any other quality assurance functions that you believe are offered by internal moderation"*). Therefore, the inferences that could be drawn from the content and number of comments made related to functions that the respondents felt had been omitted by the closed-response items, or to functions that they felt strongly enough about to choose to reiterate, emphasise, or qualify through comment. Inferences could *not* be drawn regarding the level of importance ascribed to, or perceived current emphasis on, various functions either due to inclusion in, or omission from, the qualitative data in the internal moderation and NZQA moderation survey sections.

3.4.5.7 Note about presentation of qualitative findings

When a direct quote from the data has been used in reporting the findings of this study, the respondent's organisation type has been included in parentheses at the end of each quote. The organisation type was the only required item in the survey, where all other 'demographic' information (e.g., organisation size) was collected in the last section, which not all respondents completed.

3.4.6 Stage two: Qualitative and quantitative approaches informing each other

The iterative cycle of one research approach informing another typifies mixed methods research (Teddlie & Tashakkori, 2009), as illustrated by the following three examples. Firstly, where possible, findings from the exploratory factor analysis of the quantitative survey data (3.4.4.3) were used as predetermined codes for the internal moderation and NZQA moderation qualitative data analyses. Secondly, the process of coding the internal

moderation qualitative data (3.4.5.3) precipitated a review of the interpretation of a finding of the earlier exploratory factor analysis (3.4.4.3): When the inclusions list for the predetermined factor-based code *Internal moderation: Assisting organisational development in relation to internal assessment* was consulted, consideration of the seven associated closed-response survey items revealed that the conceptual link shared by the items had been interpreted too narrowly, as reflected in the name initially given to the factor.[94] To correct this, the quantitative factor, associated scale, and associated qualitative code, were renamed to *Internal moderation: Assisting organisational development.* Thirdly, the findings of the close analysis of the *Internal moderation: Maintaining public and stakeholder confidence* qualitative data prompted further quantitative analysis (z-tests for proportions) of the relevant quantitative survey items (refer 3.4.4.10).

3.5 Methods: Summary

To conclude, this study aimed to explore the perceptions of academic leaders in three types of New Zealand education organisations about the functions of moderation. The study was based in a pragmatic paradigm and adopted a mixed methods sequential research design with multiple stages. In the first stage, semi-structured interviews were conducted with two participants from each organisation type, to inform the development of a bespoke data collection instrument for use in the second research stage. In the second stage, qualitative and quantitative data were collected via an online survey, for which a census approach had been used to invite all organisations within the target population to participate. This approach resulted in a qualitative sample of 208 respondents and a quantitative sample of 215. Quantitative and qualitative survey data were analysed separately, in parallel, but with some integration where one process informed the other. The findings of these analyses are presented in the following chapter.

---

[94] The term 'assessment' was mentioned only in item 14.3: *"To build the confidence of assessment designers or assessors"*, where the other six items did not specify the focus of development or learning.

4. Results

This chapter presents the survey results.[95] The survey sample comprised a total of 221 academic leaders, from 13 ITPs, 132 PTEs, and 76 schools (refer Table 3.1). All professed to have responsibility within their organisation for internal and NZQA moderation (see 3.4.1.1). It was not possible to ascertain how representative the sample was because not all respondents provided data that would allow this to occur (refer 3.4.1.2). Of the total sample, 215 respondents had data included in the quantitative analyses, and 208 in the qualitative analyses.

The chapter comprises four sections. Firstly, the quantitative results pertaining to internal moderation are presented, followed by the quantitative results pertaining to NZQA moderation. Next, the results of quantitative analyses comparing data about internal moderation and NZQA moderation are provided. In the last section, the qualitative results are presented. Possible interpretations of the qualitative results are discussed in relation to the literature and the context of the present study. Links are also drawn between the qualitative findings and quantitative results. Subsequent chapters contain a broader discussion in which this study's research questions are answered, the main results are compared with the literature, and suggestions are made in respect of what the findings imply for practice.

4.1 Quantitative results: Internal moderation

Academic leaders were found to perceive the functions of internal moderation as falling into the five areas of assessment quality, maintaining public confidence, organisational quality assurance, professional and organisational learning and development, and educational quality. Nine factors concerning internal moderation were revealed by quantitative analysis, which aligned with, and illuminated nuance within, these five areas.

---

[95] The results from stage one (interviews) are not presented here as the sole purpose of that stage was to inform the development of the survey instrument (see 3.3.4, 3.4.2.2.1). However, a synopsis of the interview findings is included as Appendix 4.

4.1.1 Internal moderation factors

Exploratory factor analyses were run on the closed-response survey items that addressed internal moderation, to identify any underlying constructs. These analyses resulted in the extraction of nine factors relating to the functions of internal moderation as perceived by participants. The factors are presented in an order corresponding to the main themes in the data that emerged through the stage two analysis:[96] *Checking assessor judgement quality*, *Checking assessment material quality*, *Improving assessment material and assessor judgement quality*, *Maintaining public confidence*, *Organisational quality assurance*, *Providing professional learning opportunities*, *Assisting organisational development, Checking educational quality*, and *Improving educational quality*. One item did not associate with any factor (see 4.1.2).

Table 4.1 shows the pattern matrix of internal moderation survey items associated with the *Checking assessor judgement quality* factor. This factor had an initial eigenvalue of 3.0 in the exploratory factor analysis, which accounted for 5.8% of the total variance in the data. The pattern matrix showed three items to be associated with the factor, with loading magnitudes varying from very strong to relatively weak. Item 6.4 was also associated with the factor *Improving assessment material and assessor judgement quality*, with a relatively weak loading magnitude.

The concept of checking the quality of assessor judgements is central to all three items. Two items address checking assessor judgements for accuracy and consistency. The third addresses checking that assessments are fair, to which assessor judgements contribute.

---

[96] As opposed to in descending order of the percentage of variance accounted for by each factor, as would be conventional.

Table 4.1   Pattern matrix of internal moderation items associated with the *Checking assessor judgement quality* factor.

		Percentage of variance (Extraction Sums of Squared Loadings)	
		5.8	3.2
		Factor	
Items		*Checking assessor judgement quality*	*Improving assessment material and assessor judgement quality*
Q6.5	To check that assessor judgements about assessment evidence are consistent	0.965	
Q6.4	To check that assessor judgements about assessment evidence are accurate	0.503	0.389
Q6.1	To check that assessments are fair to all students	0.398	

Note: coefficients less than 0.30 suppressed.

Figure 4.1 shows the distribution of responses for each item associated with the *Checking assessor judgement quality* factor, ordered (left to right, top to bottom) in decreasing order of magnitude of loadings on the pattern matrix. The high importance of these items was almost unanimously endorsed: Over 90% of respondents rated each as being of *High importance,* and 6% or fewer of respondents rated them as being of *Medium importance.* Furthermore, for each item the ratings of *No importance* and *Low importance* barely registered, if at all.

A principal components analysis on these three items resulted in only one component being extracted, indicating that the factor may be treated as unidimensional. However, the items would not calibrate onto a measurement scale, as there was near unanimity in the responses.

To check that assessor judgements about assessment evidence are consistent (Item 6.5; Response rate 100%; *n* = 214)

To check that assessor judgements about assessment evidence are accurate (Item 6.4; Response rate 100%; *n* = 214)

To check that assessments are fair to all students (Item 6.1; Response rate 100%; *n* = 215)
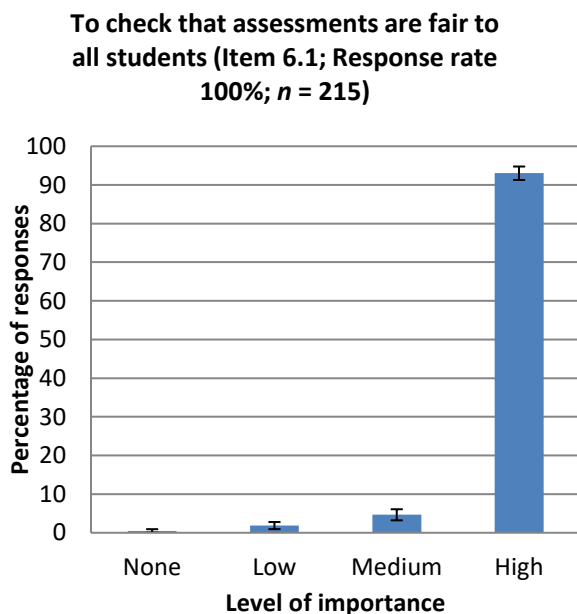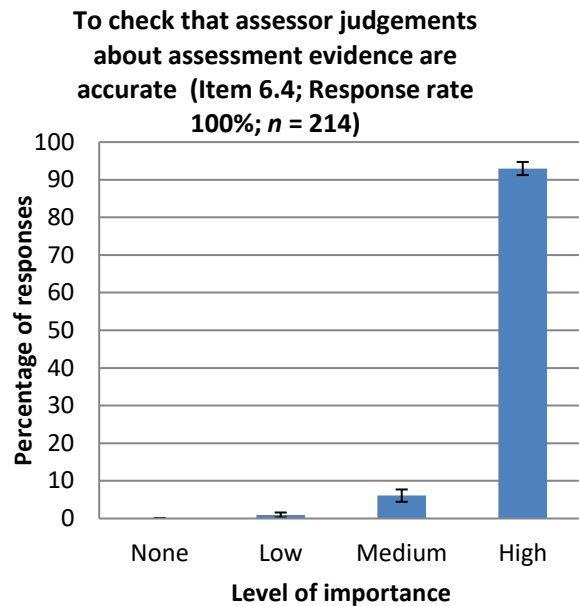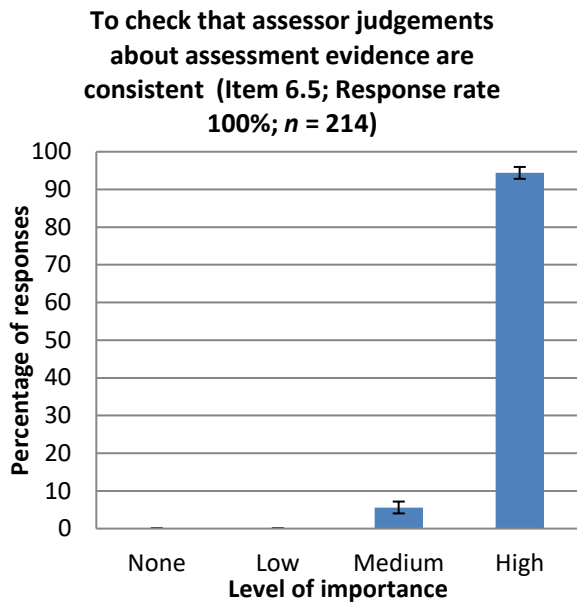
Figure 4.1   The distribution of responses for each internal moderation item contributing to the *Checking assessor judgement quality* factor. Error bars denote standard error of the percentage of responses.

The pattern matrix of internal moderation items associated with the *Checking assessment material quality* factor is shown in Table 4.2. This factor had an initial eigenvalue of 1.2 in the exploratory factor analysis, accounting for 2.4% of the total variance. Two items were shown to be strongly associated with the factor. The concept of quality control of assessment material underpins both: One item addresses checking that the content meant to be

assessed is assessed, and the other addresses checking that the assessment materials are fit-for-purpose in terms of the approved specifications or standard.

Table 4.2   Pattern matrix of internal moderation items associated with the *Checking assessment material quality* factor.

Items		Percentage of variance (Extraction Sums of Squared Loadings)
		2.3
		Factor
		Checking assessment material quality
Q6.3	To check that the assessment materials assess the content they are meant to (e.g. approved or industry-current content)	0.785
Q6.2	To check the assessment materials are fit for purpose in terms of the approved specifications or Standard	0.679

Note: coefficients less than 0.30 suppressed.

Figure 4.2 shows the distribution of responses for each item contributing to the C*hecking assessment material quality* factor. Again, there was almost unanimous endorsement of the importance of these items: Approximately 90% of respondents saw each as being of *High importance*. Most of the balance of respondents rated each as being of *Medium importance*; the rating of *Low importance* barely registered.

A principal components analysis on the two items resulted in only one component being extracted, indicating that the factor is unidimensional. However, a measurement scale could not satisfactorily be calibrated due to insufficient variability in the responses. The calibration onto a measurement scale was therefore abandoned.
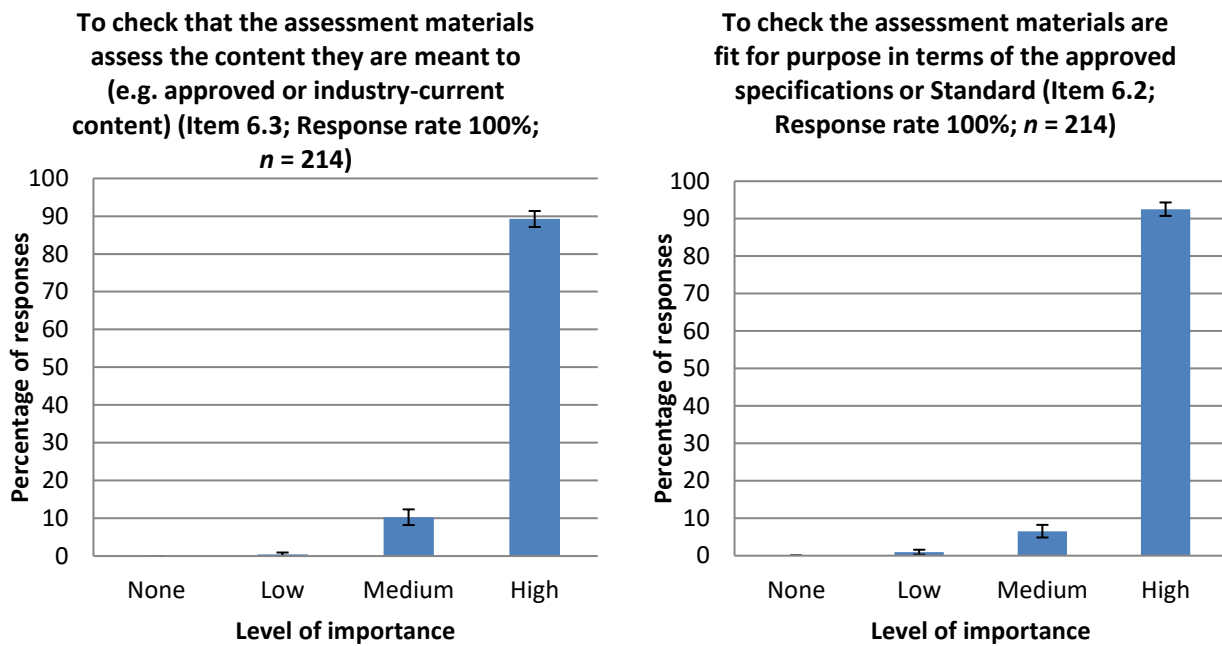
Figure 4.2 The distribution of responses for each internal moderation item contributing to the C*hecking assessment material quality* factor. Error bars denote standard error of the percentage of responses.

The pattern matrix of internal moderation items associated with the *Improving assessment material and assessor judgement quality* factor is shown in Table 4.3. This factor had an initial eigenvalue of 1.5 in the exploratory factor analysis and accounted for 3.2% of the total variance.

The four items associated with this factor share the central concept of giving feedback to improve the quality of assessment materials and assessor judgements. Two items address giving feedback about the current quality of assessor judgements and assessment materials, and the other two pertain to giving feedback about how to improve that quality.

Table 4.3   Pattern matrix of internal moderation items associated with the *Improving assessment material and assessor judgement quality* factor.

Items		Percentage of variance (Extraction Sums of Squared Loadings)	
		5.8	3.2
		Factor	
		Checking assessor judgement quality	*Improving assessment material and assessor judgement quality*
Q13.3	To give feedback about the quality of assessor judgements	0.316	0.699
Q13.4	To give feedback about how to improve assessor judgements		0.475
Q13.1	To give feedback about how well the assessment materials assess achievement against the Standard or course		0.464
Q13.2	To give feedback about how to amend the assessment materials in order to better assess achievement against the Standard or course		0.317

Note: coefficients less than 0.30 suppressed.

Figure 4.3 shows the distribution of responses for each item contributing to the *Improving assessment material and assessor judgement quality* factor. There was a clear tendency to rate each item as having *High importance*: In all cases, at least 70% of respondents rated the item as such. Far fewer respondents rated the items as being of *Medium importance* (less than 30% for item 13.1, and 20% or less for the other items). Very few respondents rated any as having *No importance* or *Low importance*.

To give feedback about the quality of assessor judgements (Item 13.3; Response rate 93%; *n* = 201)

To give feedback about how to improve assessor judgements (Item 13.4; Response rate 93%; *n* = 200)

To give feedback about how well the assessment materials assess achievement against the Standard or course (Item 13.1; Response rate 93%; *n* = 201)

To give feedback about how to amend the assessment materials in order to better assess achievement against the Standard or course (Item 13.2; Response rate 93%; *n* = 201)

Figure 4.3   The distribution of responses for each internal moderation item contributing to the *Improving assessment material and assessor judgement quality* factor. Error bars denote standard error of the percentage of responses.

The pattern matrix of internal moderation items associated with the *Maintaining public confidence* factor is shown in Table 4.4. In the exploratory factor analysis, this factor had an initial eigenvalue of 2.4, which accounted for 5.1% of the total data variance. Three survey items were strongly associated with the factor, all with strong or very strong loadings. None were associated in any substantive way with another factor.

All three items address the central concept of maintaining or ensuring public or stakeholder confidence. Specifically, the items deal with confidence in national qualifications, and in an organisation's internal assessment, and education programmes and qualifications.

Table 4.4   Pattern matrix of internal moderation items associated with the *Maintaining public confidence* factor.

Items		Percentage of variance (Extraction Sums of Squared Loadings)
		5.1
		Factor
		Maintaining public confidence
Q11.2	To maintain public confidence in the internal assessment conducted in an organisation	0.947
Q11.1	To help to ensure that other education organisations, parents, and employers can have confidence in an organisation's education programmes and qualifications	0.824
Q11.3	To maintain public confidence in national qualifications	0.707

Note: coefficients less than 0.30 suppressed.

Figure 4.4 shows the distribution of responses for each item associated with the *Maintaining public confidence* factor. There was a clear endorsement of the importance of all three items: Over half of the responses for each was in the category of *High importance*, and a further third in the category of *Medium importance*. Less than 5% of respondents rated each item as having *No importance*, and approximately 10% rated each as being of *Low importance*.

To maintain public confidence in the internal assessment conducted in an organisation (Item 11.2; Response rate 97%; *n* = 209)

To help to ensure that other education organisations, parents, and employers can have confidence in an organisation's education programmes and qualifications (Item 11.1; Response rate 97%; *n* = 209)

To maintain public confidence in national qualifications (Item 11.3; Response rate 97%; *n* = 209)

Figure 4.4 The distribution of responses for each internal moderation item contributing to the *Maintaining public confidence* factor. Error bars denote standard error of the percentage of responses.

Table 4.5 shows the pattern matrix of internal moderation items associated with the *Organisational quality assurance* factor. In the exploratory factor analysis, this factor had an initial eigenvalue of 1.9, which accounted for 4.2% of the total variance in the data. Five items were shown as being substantially associated with the factor, three of which had strong loadings, and two of which had a weaker association. The five items address various aspects of quality assurance at an organisational level: internal moderation providing information (for organisations' own quality assurance processes, or performance appraisals), being a quality assurance requirement (as per an organisation's policies, or as required by NZQA), and giving confidence to managers and governors about the quality of teaching and assessment occurring within the organisation.

Table 4.5 Pattern matrix of internal moderation items associated with the *Organisational quality assurance* factor.

Items		Percentage of variance (Extraction Sums of Squared Loadings)
		4.2
		Factor
		Organisational quality assurance
Q7.4	To provide an evidence source to be considered as part of an organisation's Self-assessment (ITPs and PTEs), or Self-review (Schools)	0.75
Q7.2	To meet the internal requirements of an organisation (e.g. as set out in the organisation's policies or quality management system)	0.746
Q7.1	To help ensure that an organisation meets NZQA's rules and requirements	0.728
Q7.6	To provide information for performance appraisals	0.427
Q7.5	To give confidence to the management team and/or governing body in the quality of teaching and assessment in their organisation	0.354

Note: coefficients less than 0.30 suppressed.

The distribution of responses for each item associated with the *Organisational quality assurance* factor is shown in Figure 4.5. There was a clear endorsement of the importance of four of the items: Approximately 70% of respondents rated each of items 7.2, 7.1 and 7.5 as being of *High importance*, as did almost 60% of the respondents for item 7.4. One quarter to one third of respondents rated each of these items as being of *Medium importance,* and less than 10% rated each as having *Low importance*.

In contrast, item 7.6 (*To provide information for performance appraisals*) attracted much less support. Forty percent of respondents rated it as having *Medium importance*, and more than 30% rated it as having *Low importance*, meaning that, in a marked difference to the rest of the internal moderation survey items, over 70% of respondents rated it as having only *Low* or *Medium importance*. This item had the lowest level of endorsement of any of the internal moderations survey items; only one fifth rated it as having *High importance*.

To provide an evidence source to be considered as part of an organisation's Self-assessment (ITPs and PTEs), or Self-review (Schools) (Item 7.4; Response rate 99%; *n* = 212)

To meet the internal requirements of an organisation (e.g. as set out in the organisation's policies or quality management system) (Item 7.2; Response rate 99%; *n* = 213)

To help ensure that an organisation meets NZQA's rules and requirements (Item 7.1; Response rate 99%; *n* = 213)

To provide information for performance appraisals (Item 7.6; Response rate 99%; *n* = 212)

To give confidence to the management team or governing body in the quality of teaching and assessment in their organisation (Item 7.5; Response rate 99%; *n* = 212)

Figure 4.5   The distribution of responses for each internal moderation item contributing to the *Organisational quality assurance* factor. Error bars denote standard error of the percentage of responses.

Table 4.6 shows the pattern matrix of internal moderation items associated with the factor, *Providing professional learning opportunities*. In the exploratory factor analysis, this factor had an initial eigenvalue of 1.3 which accounted for 2.8% of the total variance in the data. Four items were shown to be associated with the factor, three of which were associated solely with this factor (with moderate to strong loadings). The fourth item had a relatively weak loading against this factor and was also shown to have a loading of a similar magnitude against the factor, *Assisting organisational development*.

All four items address the concept that internal moderation provides opportunities for professional learning. Three address the content of the learning opportunities (learning opportunities about assessment practices, the context- or client-specific application of the requirements of the standard or course, and teaching and learning). The fourth addresses the medium though which those learning opportunities are provided (i.e., moderation feedback).

Table 4.6   Pattern matrix of internal moderation items associated with the *Providing professional learning opportunities* factor.

Items		Percentage of variance (Extraction Sums of Squared Loadings)	
		2.8	1.8
		Factor	
		Providing professional learning opportunities	*Assisting organisational development*
Q16.1	To provide learning opportunities about assessment practices	-0.843	
Q16.2	To provide learning opportunities about the context- or client-specific application of the requirements of the Standard or course	-0.581	
Q16.3	To provide learning opportunities about teaching and learning	-0.549	
Q15.1	To provide opportunities for learning through moderation feedback	-0.389	0.339

Note: coefficients less than 0.30 suppressed.

Figure 4.6 shows the distribution of responses for each item associated with the *Providing professional learning opportunities* factor. The four items were strongly supported, with a large majority of respondents indicating that each item had at least *Medium importance*. Item 15.1 (*To provide opportunities for learning through moderation feedback*) was the most

strongly endorsed of the four, with almost 70% of respondents rating it as having *High importance*, and a further almost 30% rated it as having *Medium importance.* Item 16.1 (*To provide learning opportunities about assessment* practices) was also strongly endorsed: Almost 60% rated it as having *High importance*, and approximately one third rated it as having *Medium importance*.

To provide learning opportunities about assessment practices (Item 16.1; Response rate 92%; *n* = 197)

To provide learning opportunities about the context- or client-specific application of the requirements of the Standard or course (Item 16.2; Response rate 92%; *n* = 198)

To provide learning opportunities about teaching and learning (Item 16.3; Response rate 92%; *n* = 198)

To provide opportunities for learning through moderation feedback (Item 15.1; Response rate 92%; *n* = 197)

Figure 4.6 The distribution of responses for each internal moderation item contributing to the *Providing professional learning opportunities* factor. Error bars denote standard error of the percentage of responses.

The pattern matrix of internal moderation items associated with the *Assisting organisational development* factor is shown in Table 4.7. This factor had an initial eigenvalue of 1.1 in the exploratory factor analysis, which accounted for 1.8% of the total variance in the data. The pattern matrix showed seven items as being associated with the factor, with relatively strong to relatively weak loading magnitudes. Two of the items (15.2 and 15.3) were shown to also be associated with the factor, *Providing professional learning opportunities.* The other five were not substantially associated with any other factor.

The items associated with this factor all address the concept of assisting organisational development. Five address assisting with personnel development: building people's confidence, assisting people to develop a shared understanding, providing opportunities for professional or collegial conversations, providing guidance about how to interpret the requirements of a course or standard, and providing learning opportunities through one-on-one support. The other two inform the organisation of professional development needs, and where quality assurance activities should be focused.

Table 4.7   Pattern matrix of internal moderation items associated with the *Assisting organisational development* factor.

		Percentage of variance (Extraction Sums of Squared Loadings)	
		2.8	1.8
		Factor	
Items		*Providing professional learning opportunities*	*Assisting organisational development*
Q14.3	To build the confidence of assessment designers or assessors		0.675
Q14.2	To assist in the development of a shared understanding among relevant people		0.629
Q14.4	To alert the organisation to where professional development is required		0.598
Q15.2	To provide opportunities for professional / collegial conversations	-0.317	0.543
Q14.1	To give guidance about how to interpret the requirements of the Standard or course		0.458
Q15.3	To provide opportunities for learning through one-on-one support	-0.418	0.43
Q8.3	To show where more quality assurance activity should be focused within an organisation		0.372

Note: coefficients less than 0.30 suppressed.

Figure 4.7 shows the distribution of responses for each item associated with the *Assisting organisational development* factor. Respondents showed strong support for the functions addressed by the seven items: More than 85% of respondents rated each item as being of at least *Medium importance*, and for four items, more than 90% of respondents did.
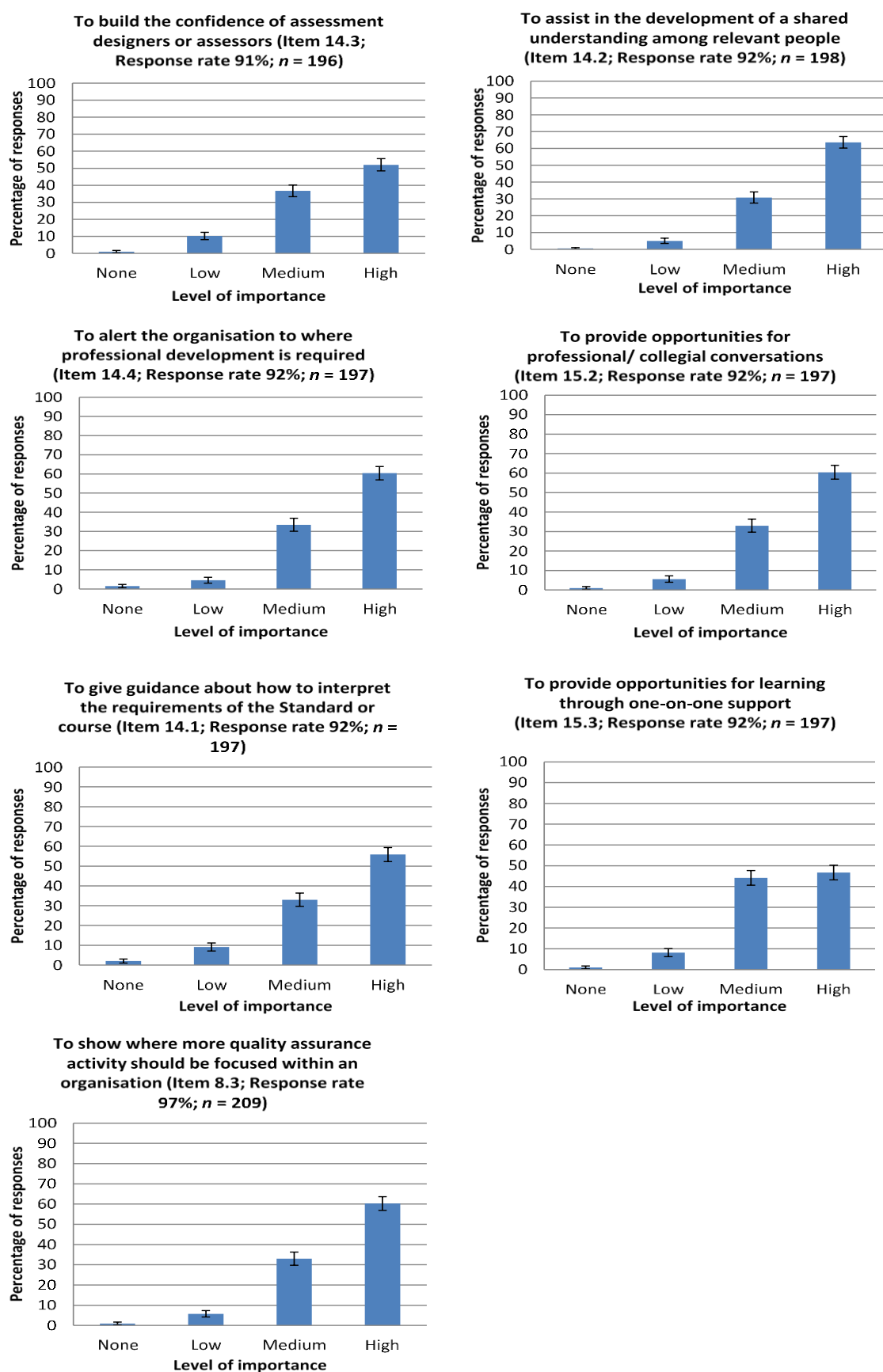
Figure 4.7   The distribution of responses for each internal moderation item contributing to the *Assisting organisational development* factor. Error bars denote standard error of the percentage of responses.

Table 4.8 presents the pattern matrix of the internal moderation items associated with the *Checking educational quality* factor. This factor had an initial eigenvalue of 2.0 in the exploratory factor analysis, accounting for 4.9% of the total variance. Five items were shown to be substantially associated with the factor by the pattern matrix: Three had strong loading magnitudes, and two had weaker loadings onto the factor. All five items focus on checking that certain aspects of education are of an acceptable quality: student achievement levels, how well students are prepared for further education or employment, how well teaching has prepared students for assessment, the currency of the curriculum, and the consistency of students' assessment experience.

Table 4.8   Pattern matrix of internal moderation items associated with the *Checking educational quality* factor.

		Percentage of variance (Extraction Sums of Squared Loadings)
		4.9
		Factor
Items		*Checking educational quality*
Q9.2	To check that students are achieving at the levels that the organisation deems they should be achieving at	0.813
Q9.3	To check that students are well prepared for going into further education or employment	0.798
Q9.1	To check that the teaching that students have received has adequately prepared them for assessment	0.766
Q8.4	To check that the approved and taught curriculum is current	0.439
Q8.2	To check the consistency of students' assessment experience throughout a programme	0.360

Note: coefficients less than 0.30 suppressed.

The distribution of responses for each item associated with the C*hecking educational quality* factor is shown in Figure 4.8. There was a clear endorsement of the importance of the five items: A distinct majority of respondents (two thirds or more) rated each as having at least *Medium importance*. The strongest support was for item 8.2 (*To check the consistency of students' assessment experience throughout a programme*), with two thirds of respondents rating it as having *High importance,* and a further one quarter rating it as being of *Medium importance*.
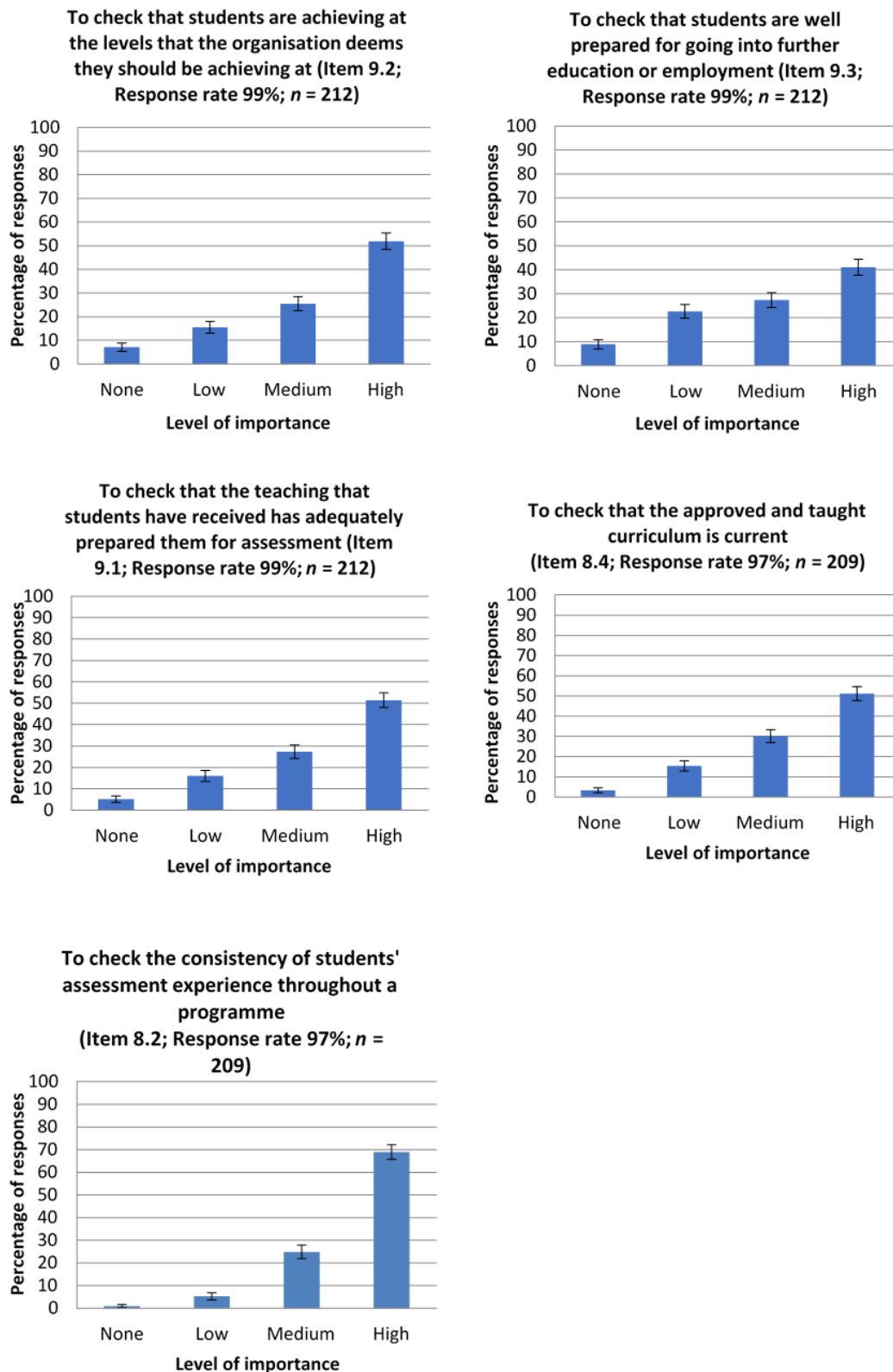
Figure 4.8 The distribution of responses for each internal moderation item contributing to the *Checking educational quality* factor. Error bars denote standard error of the percentage of responses.

Table 4.9 shows the pattern matrix of internal moderation items associated with the *Improving educational quality* factor. In the exploratory factor analysis, this factor had an initial eigenvalue of 11.9, accounting for 29.5% of the total variance. The pattern matrix showed four survey items strongly associated with this factor. None were substantially associated with any other factor. All four items contributing to this factor address informing the improvement of various aspects of education (teaching, curriculum review, assessment review, and reviews of student performance).

Table 4.9 Pattern matrix of internal moderation items associated with the *Improving educational quality* factor.

Items		Percentage of variance (Extraction Sums of Squared Loadings)
		29.5
		Factor
		Improving educational quality
Q18.2	To inform teaching	0.751
Q18.1	To inform curriculum reviews (e.g. its currency, relevance, or tailoring for particular students)	0.712
Q18.3	To inform assessment review	0.631
Q18.4	To inform reviews of student performance	0.627

Note: coefficients less than 0.30 suppressed.

Figure 4.9 shows the distribution of responses for each item contributing to the *Improving educational quality* factor. A clear majority of respondents endorsed the importance of these items, with over 70% rating each as being of *Medium* or *High importance.* Items 18.2 (*To inform teaching*) and 18.3 (*To inform assessment review*) were the most strongly endorsed of the items, with over half of the respondents rating each as being of *High importance*, and approximately 30% rating them as being of *Medium importance.*
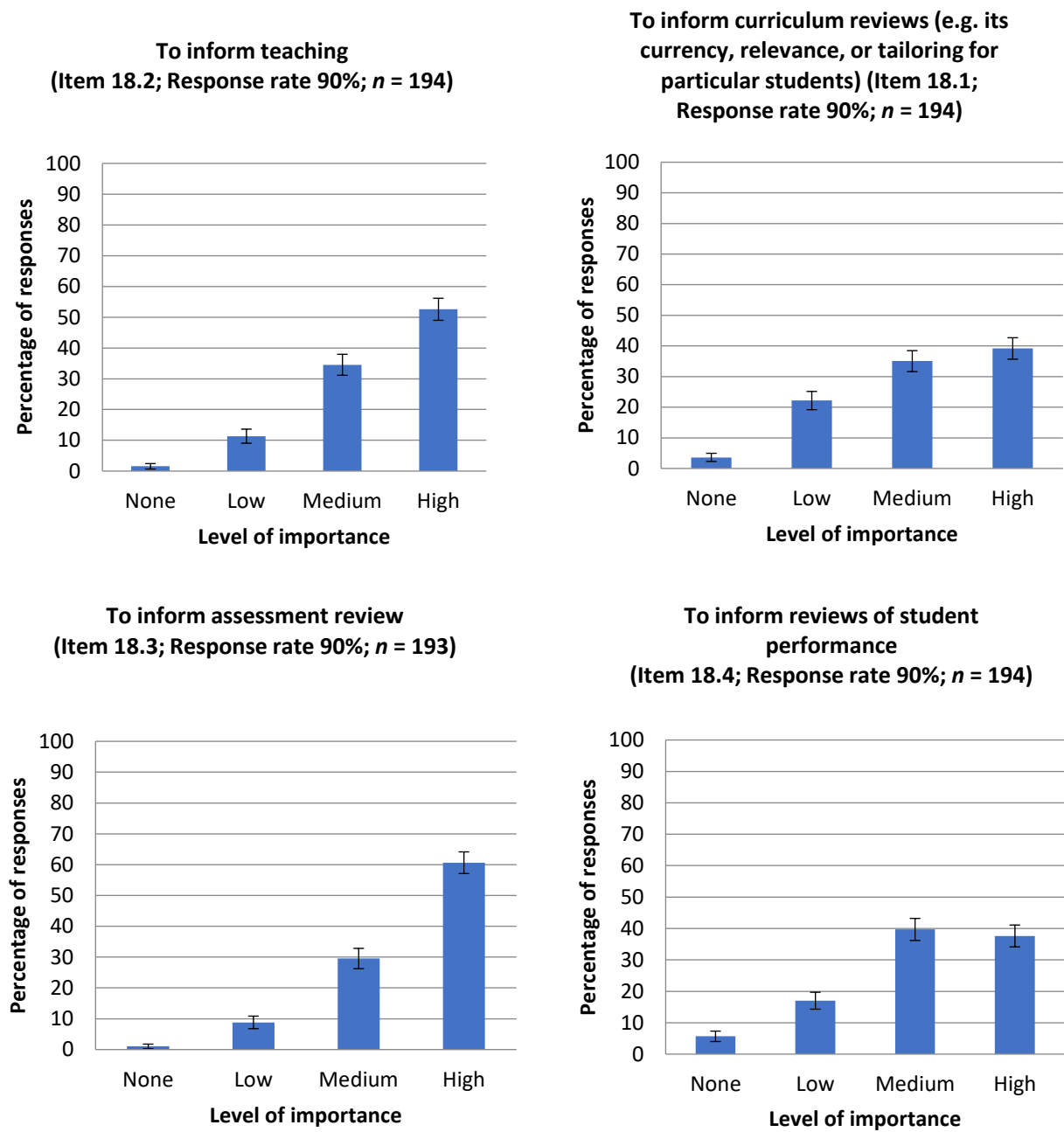
Figure 4.9 The distribution of responses for each internal moderation item contributing to the *Improving educational quality* factor. Error bars denote standard error of the percentage of responses.

4.1.2 Item that did not load onto a factor

In the exploratory factor analysis of internal moderation survey items, item 8.1: *To check that the assessment practices in an organisation are robust*, (response rate 97%, $n = 209$), did not load onto a factor in the pattern or structure matrices. There was near unanimity in respondents' rating for this item, with 90% rating it as having *High importance*, and the remainder rating it as having *Medium importance*.

4.1.3 Measurement scale mean locations of organisation types

Figure 4.10 shows the mean measurement scale locations for each of the three organisation types on each of the internal moderation measurement scales. The figure suggests variation between the mean locations of the three organisation types on each scale.

On three scales, *Organisational quality assurance*, *Checking educational quality*, and *Improving educational quality*, PTEs showed by far the greatest propensity to endorse the associated items, while schools showed the lowest propensity to endorse them. ITPs showed only slightly more of a propensity to endorse the items relating to the *Organisational quality assurance* and *Checking educational quality* scales than schools did.

These propensities were borne out in one-way analyses of variance and independent samples *t*-tests that were conducted on the scales. For each, significant differences were found between the mean locations for PTEs and schools, and for two, also between PTEs and ITPs.[97]

On the *Organisational quality* assurance scale, a one-way analysis of variance showed significant difference amongst the means of the three organisation types: $F_{(2, 212)} = 15.28$, $p < .001$. Independent samples *t*-tests revealed that the difference between the PTE and school means was significant: $t_{(198)} = 5.29$, $p < .001$, as was the difference between the PTE and ITP means: $t_{(137)} = 2.16$, $p = .03$. No significant difference was found between the ITP and school means: $t < 1$.

On the *Checking educational quality* scale, a one-way analysis of variance showed significant difference amongst the three means: $F_{(2, 210)} = 29.60$, $p < .001$. Independent

---

[97] Throughout, the low statistical power of the *t*-tests involving ITPs is noted, due to the small ITP sample size.

sample *t*-tests showed that the difference between PTEs and schools was significant: t (196) = 7.60, p < .001, as was the difference between PTEs and ITPs: t (136) = 2.60, p = .01. There was no significant difference between the means of ITPs and schools: t (84) = 1.27, p = .21.

The one-way analysis of variance on the *Improving educational quality* scale showed significant difference amongst the three means: F (2, 193) = 5.99, p = .003. Independent samples *t*-tests showed that the difference between the means for PTEs and schools was significant: t (179) = 3.42, p = .001. However, there was no significant difference between the means for ITPs and either of the two other organisation types: ITP and PTE: t < 1; ITP and school: t (81) = 1.18, p = .24.

In contrast, one-way analyses of variance showed that there were no significant differences amongst the mean locations of the three organisation types for *Maintaining public confidence*: F (2, 205) = 1.34, p = .27; *Improving assessment material and assessor judgement quality*: F (2, 200) = 1.83, p = .16; *Providing professional learning opportunities*: F (2, 194) = 2.18, p = .12; or *Assisting organisational development*: F < 1.
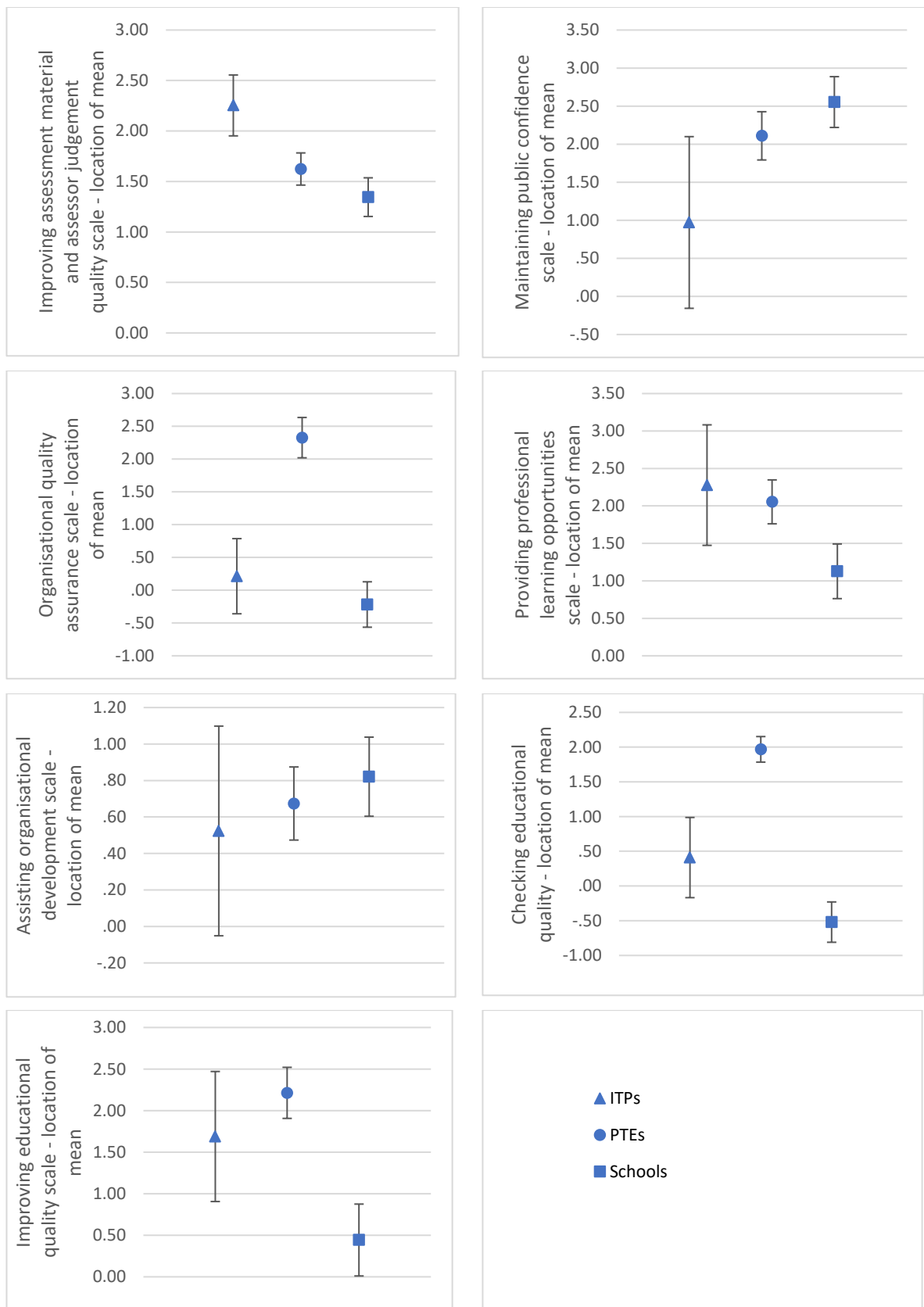
Figure 4.10   Measurement scale location of means for ITP, PTE and school respondents on each internal moderation scale. Error bars denote standard errors of the means.

4.1.4 Other tests

4.1.4.1 Pearson's correlations between internal moderation scales

Pearson's correlation coefficients between internal moderation scales for the aggregated organisation data and for the individual organisation types are shown in Table 4.10. The correlations between the internal moderation scales for the aggregated organisations were all positive and significant at the 0.01 level, although they varied in strength. However, the correlations between scales according to individual organisation types varied considerably, both in strength and statistical significance. The correlations between scales tended to be stronger for PTEs than the corresponding correlations were for ITPs or schools.

Table 4.10 Pearson's correlation coefficients between internal moderation scales, for aggregated organisations, and for individual organisation types

Internal moderation scale	Organisation type	Internal moderation scale					
		4	5	6	7	8	9
3 Improving assessment material and assessor judgement quality	Aggregate	**.37**	**.34**	**.31**	**.48**	**.32**	**.41**
	ITPs						
	PTEs	**.43**	**.36**	**.30**	**.54**	**.42**	**.42**
	Schools	**.31**	.30	.30	**.45**	.26	**.41**
4 Maintaining public confidence	Aggregate		**.42**	**.37**	**.45**	**.30**	**.37**
	ITPs						
	PTEs		**.50**	**.44**	**.52**	**.53**	**.50**
	Schools		**.43**	**.33**	**.36**		.31
5 Organisational quality assurance	Aggregate			**.50**	**.50**	**.59**	**.56**
	ITPs						
	PTEs			**.54**	**.59**	**.56**	**.50**
	Schools			**.45**	**.51**	**.48**	**.58**
6 Providing professional learning opportunities	Aggregate				**.74**	**.49**	**.61**
	ITPs				**.93**		.68
	PTEs				**.79**	**.56**	**.61**
	Schools				**.64**	**.40**	**.56**
7 Assisting organisational development	Aggregate					**.42**	**.60**
	ITPs						**.73**
	PTEs					**.58**	**.68**
	Schools					**.39**	**.52**
8 Checking educational quality	Aggregate						**.67**
	ITPs						
	PTEs						**.69**
	Schools						**.62**
9 Improving educational quality							

Note: Correlations in boldface are significant at the 0.01 level (2-tailed), and correlations in standard font are significant at the 0.05 level (2-tailed). Non-significant correlations have been omitted.

The strongest correlation for the aggregated data was between the scales, *Providing professional learning opportunities* and *Assisting organisational development*, ($r = .74$). This is reflected in the strength of the correlation for each of the three organisation types: for schools, $r = .64$; for PTEs, $r = .79$; and for ITPs, r = .93, an almost perfect correlation.[98] This

---

[98] It is acknowledged that correlations are suppressed when endorsement rates are very high, due to a lack of variability in the data.

is one of only three correlations between internal moderation scales for ITPs that were statistically significant.[99]

For the aggregated data, the *Improving educational quality* scale was strongly correlated with most other internal moderation scales: with *Checking educational quality* ($r = .67$), *Providing professional learning opportunities* ($r = .61$), *Assisting organisational development* ($r = .60$), and *Organisational quality assurance* ($r = .56$). The magnitude of three of these correlations was greater for PTEs than for schools (e.g., with *Checking educational quality*: $r = .69$ for PTES, and $r = .62$ for schools). The *Improving educational quality* scale was more strongly correlated with *Organisational quality assurance* for schools ($r = .58$) than for PTEs ($r = .50$). For ITPs, the scale, *Improving educational quality* was strongly correlated with two internal moderation scales: *Providing professional learning opportunities*, ($r = .68$), and *Assisting organisational development*, ($r = .73$).

The correlations between the *Improving educational quality* scale and two other internal moderation scales are much stronger for PTEs than for schools: *Assisting organisational development*, for PTEs, $r = .68$, whereas for schools, $r = .52$; *Maintaining public confidence* scale, for PTEs, $r = .50$, whereas for schools, $r = .31$.

The *Organisational quality assurance* scale was also strongly correlated with three other internal moderation scales for the aggregated data (*Checking educational quality*, *Providing professional learning opportunities,* and *Assisting organisational development* scales). All correlations bar one between the *Organisational quality assurance* scale and other internal moderation scales for PTEs were strong. For schools, the *Organisational quality assurance* scale had moderate or strong correlations with other internal moderation scales.

For the aggregated data, the *Maintaining public confidence* and *Improving assessment material and assessor judgement quality* scales were only moderately correlated with other internal moderation scales. The weakest correlation for the aggregated organisations was between the *Maintaining public confidence* and *Checking educational quality* scales. The correlation between these two scales for ITPs and schools was non-significant, while, in marked contrast, the corresponding correlation for PTEs was strong and significant. Further, for PTEs, the *Maintaining public confidence* scale was also strongly correlated with three other internal moderation scales: *Assisting organisational development*, *Improving*

---

[99] Although the low statistical power of the correlation coefficients for ITPs due to the small sample size is noted.

educational quality, and *Organisational quality assurance.* For schools, the corresponding correlations were only moderate in strength, and in the case of one, non-significant.

For PTEs, only the correlation between the *Improving assessment material and assessor judgement quality* scale *,* and the *Assisting organisational development* scale was strong; all other correlations involving the *Improving assessment material and assessor judgement quality* scale were moderate. The corresponding correlations for schools were all moderate, or in one case, weak.

4.1.4.2 Correlations between internal moderation scales, and length of service

Pearson's correlation coefficients were calculated between each of the internal moderation scales and respondents' length of service as a teacher, assessor, or administrator in formal education. The only statistically significant correlation (2-tailed) found was very weak and negative—the *Improving educational quality* scale was weakly and negatively correlated with respondents' length of service: $r = -.16$, $p = .04$.

Pearson's correlation coefficients were also calculated between each of the internal moderation scales and respondents' length of service in academic leadership and management roles. No statistically significant correlations were found.

4.1.4.3 Chi-square tests for independence

Chi-square tests for independence were conducted on those items contributing to the two factors that could not be calibrated (*Checking assessor judgement quality*, and *Checking assessment material quality*), and the item that did not load against any of the factors. The chi-square tests for independence on each of the six items (6.1, 6.2, 6.3, 6.4, 6.5, and 8.1) were discarded. In each there was only one degree of freedom, reflecting that the responses were aggregated across the two neighbouring categories of *Medium importance* and *High importance* (as seen in Figures 4.1 and 4.2, and in 4.1.2).

4.1.5 Internal moderation quantitative results: Summary

In summary, the quantitative analyses suggest that the factors describing academic leaders' perceptions of the functions of internal moderation aligned with the five main areas of assessment quality, maintaining public and stakeholder confidence, organisational quality assurance, professional and organisational learning and development, and educational quality. Academic leaders in all three organisation types generally saw all functions of internal moderation as important. The level of importance ascribed to individual survey items varied from almost unanimous endorsement of the high importance of those items concerning the checking of assessment quality, to about two thirds of respondents rating items associated with education quality as having at least medium importance. Correlations between the internal moderation scales showed that, for the most part, people responded about the assessment quality scale in a way that was only weakly or moderately aligned with how they responded to other scales. In contrast, there was a group of broader functions that are not directly about assessment quality (e.g., educational quality) that 'hung together'—that is, that people tended to respond similarly to.

Respondents from PTEs tended to rate organisational quality assurance and educational quality functions as being of higher importance than those from schools (and for two of the three scales, than ITPs). There was also a marked difference between PTEs and schools in how strongly some of the broader (non-assessment quality) functions hung together (i.e., the strength of correlations), with leaders from PTEs tending to respond to those functions in a similar way more so than schools.

4.2 Quantitative results: NZQA moderation

Academic leaders' perceptions of the functions of NZQA moderation were found to fall into four areas: assessment quality, maintaining public confidence, organisational quality assurance, and professional and organisational learning and development. Quantitative analysis revealed five factors concerning NZQA moderation which aligned with, and illuminated nuance within, these four areas. The factors, *Checking internal assessment quality* and *Improving internal assessment quality,* aligned with the assessment quality area; the *Maintaining public confidence* factor aligned with the maintaining public confidence area; the *Organisational quality assurance* factor aligned with the organisational quality assurance area; and the factor, *Providing opportunities to learn from NZQA moderation reports*, aligned with professional and organisational learning and development.

4.2.1 NZQA moderation factors

Exploratory factor analyses were conducted on the closed-response survey items pertaining to the functions of NZQA moderation, resulting in five factors being extracted. These factors are presented in an order that corresponds to the main themes in the data:[100] *Checking internal assessment quality, Improving internal assessment quality*, *Maintaining public confidence*, *Organisational quality assurance*, and *Providing opportunities to learn from NZQA moderation reports*. Two survey items were not substantially associated with any factor (see 4.2.2).

Table 4.11 shows the pattern matrix of NZQA moderation items associated with the *Checking internal assessment quality* factor. This factor had an initial eigenvalue of 1.1 in the exploratory factor analysis and accounted for 5.2% of the total variance. The pattern matrix showed that three survey items were associated with this factor, one with a strong loading magnitude, and two with moderate loading magnitudes. None was substantially associated with any other factors.

The three items address checking the quality of internal assessments: checking the accuracy of assessor judgements against the requirements of the standard, that assessment materials are fit-for-purpose in terms of those requirements, and that an organisation's internal assessments meet those requirements.

---

[100] Instead of as would conventionally be presented (in descending order of the percentage of variance accounted for by each factor).

Table 4.11   Pattern matrix of NZQA moderation items associated with the *Checking internal assessment quality* factor.

Items		Percentage of variance (Extraction Sums of Squared Loadings)
		5.2
		Factor
		Checking internal assessment quality
Q21.3	To check that assessor judgements are accurate against the requirements of the Standard	.764
Q21.2	To check that assessment materials are fit for purpose in terms of the requirements of the Standard	.617
Q21.1	To check whether or not the internal assessments of an organisation meet the requirements of the Standards, as set nationally	.493

Note: coefficients less than 0.30 suppressed.

Figure 4.11 shows the distribution of responses for each item associated with the C*hecking internal assessment quality* factor. Respondents perceived that NZQA moderation had the greatest emphasis on these items of any of the NZQA moderation items. There was almost unanimous agreement with the rating of *High emphasis* across them, with 80-90% of respondents rating each item as such. For each, only around 10% of respondents indicated that there was *Medium emphasis* on the item, while the categories of *No emphasis* and *Low emphasis* barely registered.

To check that assessor judgements are accurate against the requirements of the Standard (Item 21.3; Response rate 87%; *n* = 188)
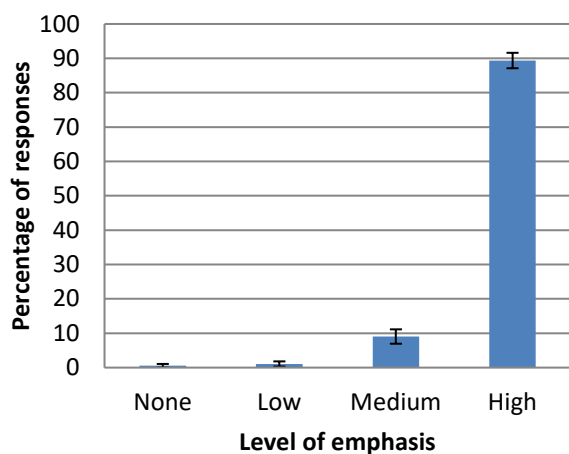
To check that assessment materials are fit for purpose in terms of the requirements of the Standard (Item 21.2; Response rate 87%; *n* = 188)
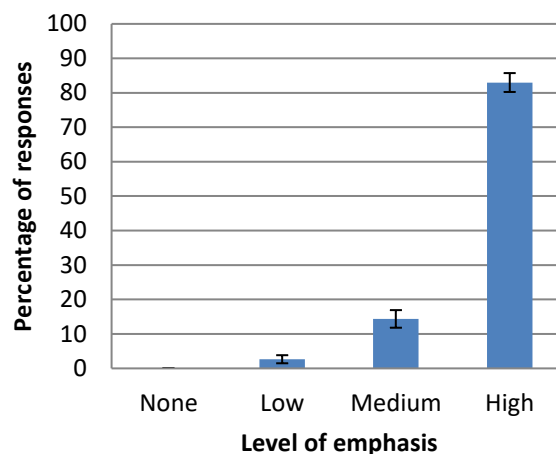
To check whether or not the internal assessments of an organisation meet the requirements of the Standards, as set nationally (Item 21.1; Response rate 87%; *n* = 188)
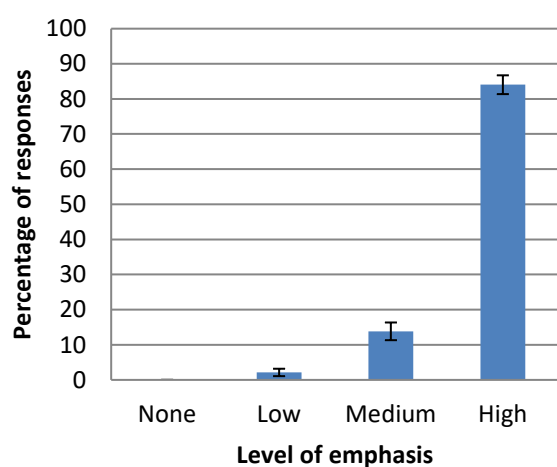
Figure 4.11   The distribution of responses for each NZQA moderation item contributing to the *Checking internal assessment quality* factor. Error bars denote standard error of the percentage of responses.

Table 4.12 shows the pattern matrix of NZQA moderation items associated with the factor, *Improving internal assessment quality*. In the exploratory factor analysis, this factor had an initial eigenvalue of 2.0, accounting for 7.2% of the total variance. The pattern matrix showed eight survey items associated with this factor, six of which had strong loading magnitudes. The two other items also were weakly associated with other factors: Item 26.5 (*To give feedback about the quality of the assessor judgements*) was also associated with the factor, *Checking internal assessment quality*, and item 29.2 (*To give feedback about whether an assessment supports teaching and learning*) was also associated with the factor, *Organisational quality assurance*.

The objective of improving the quality of internal assessment underpins all eight items, with each item addressing a different facet of this function. Four items address the development of personnel in relation to internal assessment: developing a shared understanding, stimulating professional or collegial conversations, building the confidence of assessment designers and assessors, and giving guidance about how to interpret the standard requirements. Two items focus on improving the quality of assessor judgements, and one addresses how to improve the assessment material. The final item considers whether an assessment supports teaching and learning.

Table 4.12   Pattern matrix of NZQA moderation items associated with the *Improving internal assessment quality* factor.

		Percentage of variance (Extraction Sums of Squared Loadings)		
		7.2	8.8	5.2
		Factor		
Items		Improving internal assessment quality	Organisational quality assurance	Checking internal assessment quality
Q27.3	To assist in the development of a shared understanding among relevant people	.784		
Q27.2	To stimulate professional and collegial conversation	.744		
Q27.4	To build the confidence of assessment designers or assessors	.735		
Q27.1	To give guidance regarding how to interpret the requirements of a Standard	.708		
Q26.6	To give feedback about how to improve assessor judgements	.605		
Q26.4	To give feedback about how to amend the assessment materials in order to better assess achievement against a Standard	.524		
Q26.5	To give feedback about the quality of the assessor judgements	.509		.380
Q29.2	To give feedback about whether an assessment supports teaching and learning	.388	.351	

Note: coefficients less than 0.30 suppressed.

Figure 4.12 shows the distribution of responses for each item associated with the factor *Improving internal assessment quality*. Over 70% of respondents considered there to be at least *Medium emphasis* on three of the items, as did over 60% for three other items and over half of the respondents for a seventh item. The question asked of the last item associated with this factor (item 29.2) was regarding its importance as a function of NZQA moderation. This item was strongly endorsed, with over 70% of respondents rating it as having at least *Medium importance.* Less than 10% of respondents indicated that any item associated with this factor received *No emphasis* or had *No importance*. However, there was lower perceived emphasis placed on of the four items addressing building professional confidence and shared understandings (27.1, 27.2, 27.3, 27.4) than the other survey items pertaining to NZQA moderation.
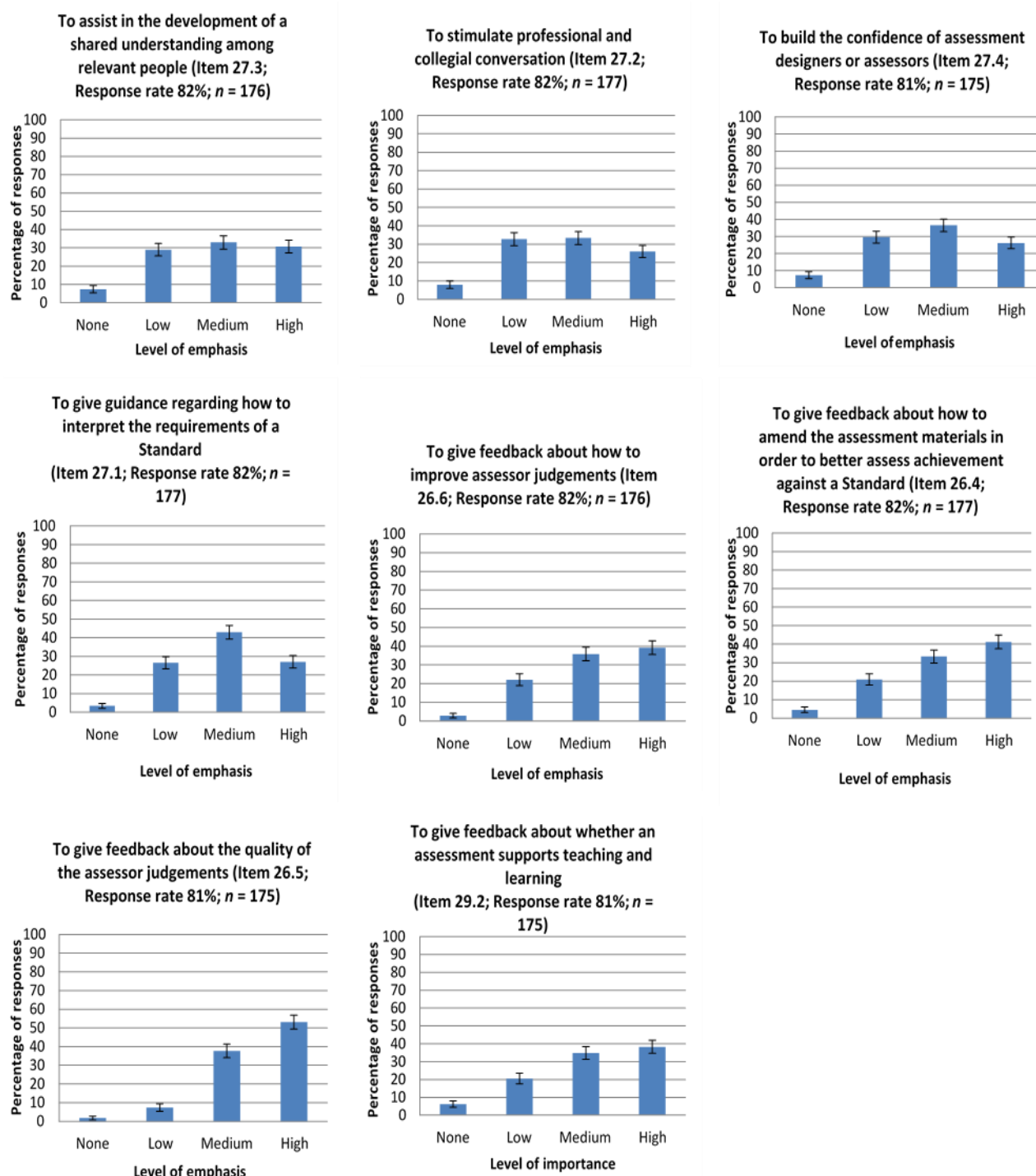
Figure 4.12 The distribution of responses for each NZQA moderation item contributing to the *Improving internal assessment quality* factor. Error bars denote standard error of the percentage of responses.

The pattern matrix of the three NZQA moderation items associated with the *Maintaining public confidence* factor is shown in Table 4.13. This factor had an initial eigenvalue of 2.7 in the exploratory factor analysis and accounted for 11.8% of the total variance. The pattern matrix showed that three items were strongly associated with this factor, and none was substantially associated with any other factors.

The maintenance of public confidence is central to all three items. Two explicitly address maintaining public confidence: in the internal assessments conducted by organisation, and in national qualifications. The third addresses ensuring that the education and qualifications provided are seen as credible by stakeholders, which is requisite to public confidence being maintained.

Table 4.13   Pattern matrix of NZQA moderation items associated with the *Maintaining public confidence* factor.

Items		Percentage of variance (Extraction Sums of Squared Loadings)
		11.8
		Factor
		Maintaining public confidence
Q24.2	To maintain public confidence in the internal assessment conducted by organisations	.888
Q24.1	To help to ensure that education and qualifications are seen as credible by parents, employers and other education organisations	.873
Q24.3	To maintain public confidence in national qualifications	.737

Note: coefficients less than 0.30 suppressed.

Figure 4.13 shows the distribution of responses for each item associated with the *Maintaining public confidence* factor. There was a strong perceived emphasis by NZQA moderation on the functions addressed by these items, with more than 60% of respondents rating each as receiving *High emphasis*, and a full 90% of respondents rating each as receiving at least *Medium emphasis*.

To maintain public confidence in the internal assessment conducted by organisations
(Item 24.2; Response rate 87%; *n* = 187)

To help to ensure that education and qualifications are seen as credible by parents, employers and other education organisations (Item 24.1; Response rate 87%; *n* = 187)

To maintain public confidence in national qualifications (Item 24.3; Response rate 87%; *n* = 187)
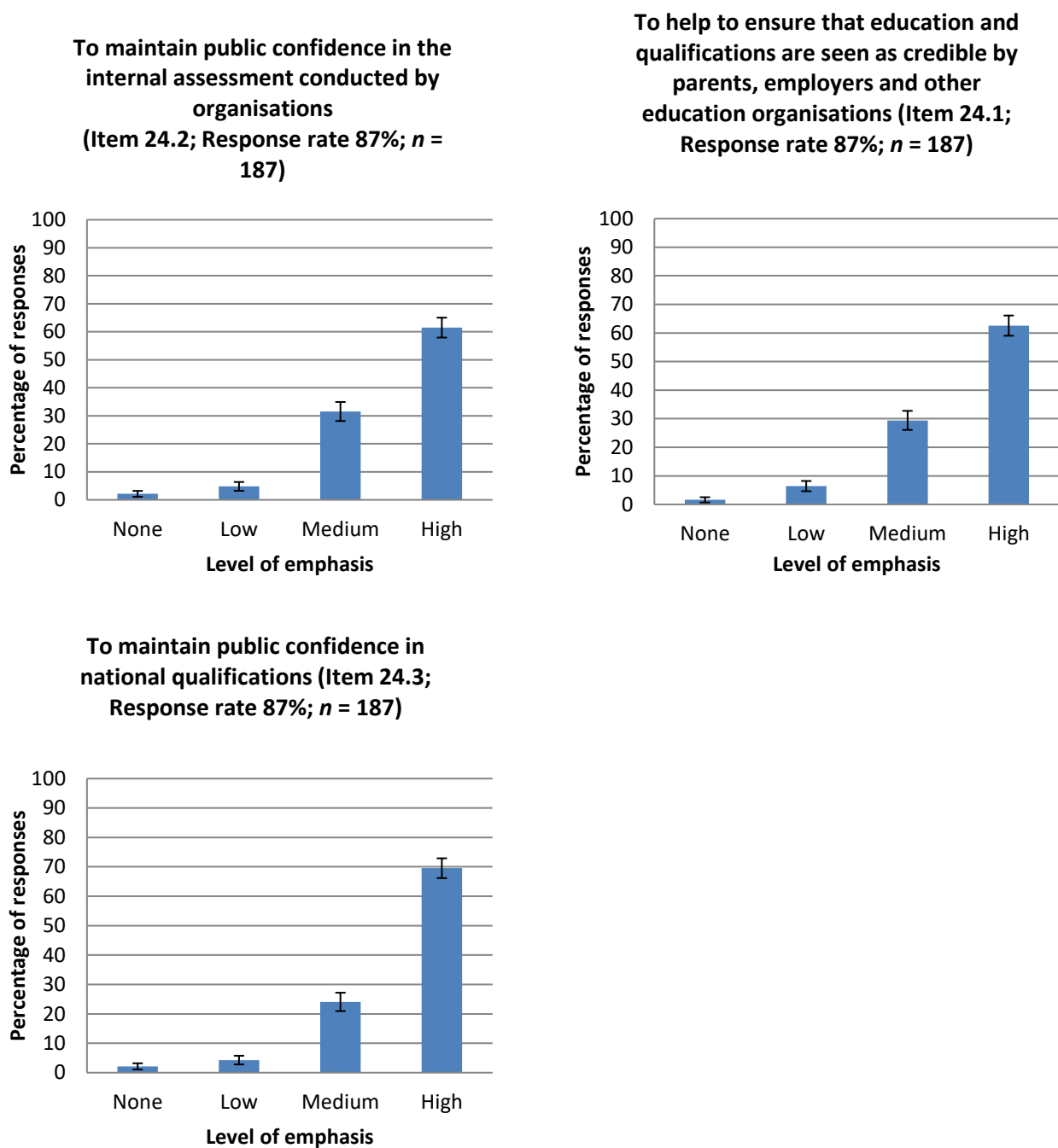
Figure 4.13 The distribution of responses for each NZQA moderation item contributing to the *Maintaining public confidence* factor. Error bars denote standard error of the percentage of responses.

Table 4.14 shows the pattern matrix of NZQA moderation items associated with the *Organisational quality assurance* factor. This factor had an initial eigenvalue of 1.9 in the exploratory factor analysis and accounted for 8.8% of the total variance. The pattern matrix showed seven survey items to be associated with this factor, one with a strong loading, and

the rest with moderate to strong loadings. None were associated in any substantial way with another factor.

These seven items address functions involved with quality assurance at an organisation level: providing information (for overarching internal and external quality assurance processes, and performance appraisals), satisfying externally imposed requirements, and maintaining the confidence of managers or the governing body. The functions also include ascertaining whether appropriate content is assessed by an organisation, and whether those assessments are fair to all students.

Table 4.14   Pattern matrix of NZQA moderation items associated with the *Organisational quality assurance* factor.

		Percentage of variance (Extraction Sums of Squared Loadings)
		8.8
		Factor
Items		*Organisational quality assurance*
Q22.1	To provide evidence for consideration in an organisation's Self-assessment (ITPs and PTEs), or Self-review (Schools)	.765
Q22.2	To provide an evidence source that contributes to an organisation's External Evaluation and Review, and Provider Category (ITPs and PTEs), or Managing National Assessment review (Schools)	.643
Q22.3	To monitor an organisation's compliance with NZQA's rules and requirements	.622
Q22.5	To provide information for performance appraisals	.581
Q22.4	To give confidence to the management team or governing body about an organisation's teaching and assessment	.567
Q21.5	To provide information about whether or not content that is assessed is appropriate	.535
Q21.4	To check that assessments are fair to all students	.492

Note: coefficients less than 0.30 suppressed.

Figure 4.14 shows the distribution of responses for each item contributing to the *Organisational quality assurance* factor. A clear majority of respondents believed that NZQA moderation places weight on each item (except for item 22.5): Over half considered that each received *High emphasis*, and 85% or more rated each as receiving at least *Medium emphasis*.

In stark contrast, respondents felt that NZQA moderation placed the weakest emphasis of any of the NZQA moderation survey items on item 22.5 (*To provide information for performance appraisals*). One half of respondents rated it as receiving either *No emphasis* or only *Low emphasis*. The remaining respondents were equally split between the ratings of *Medium* and *High emphasis.*
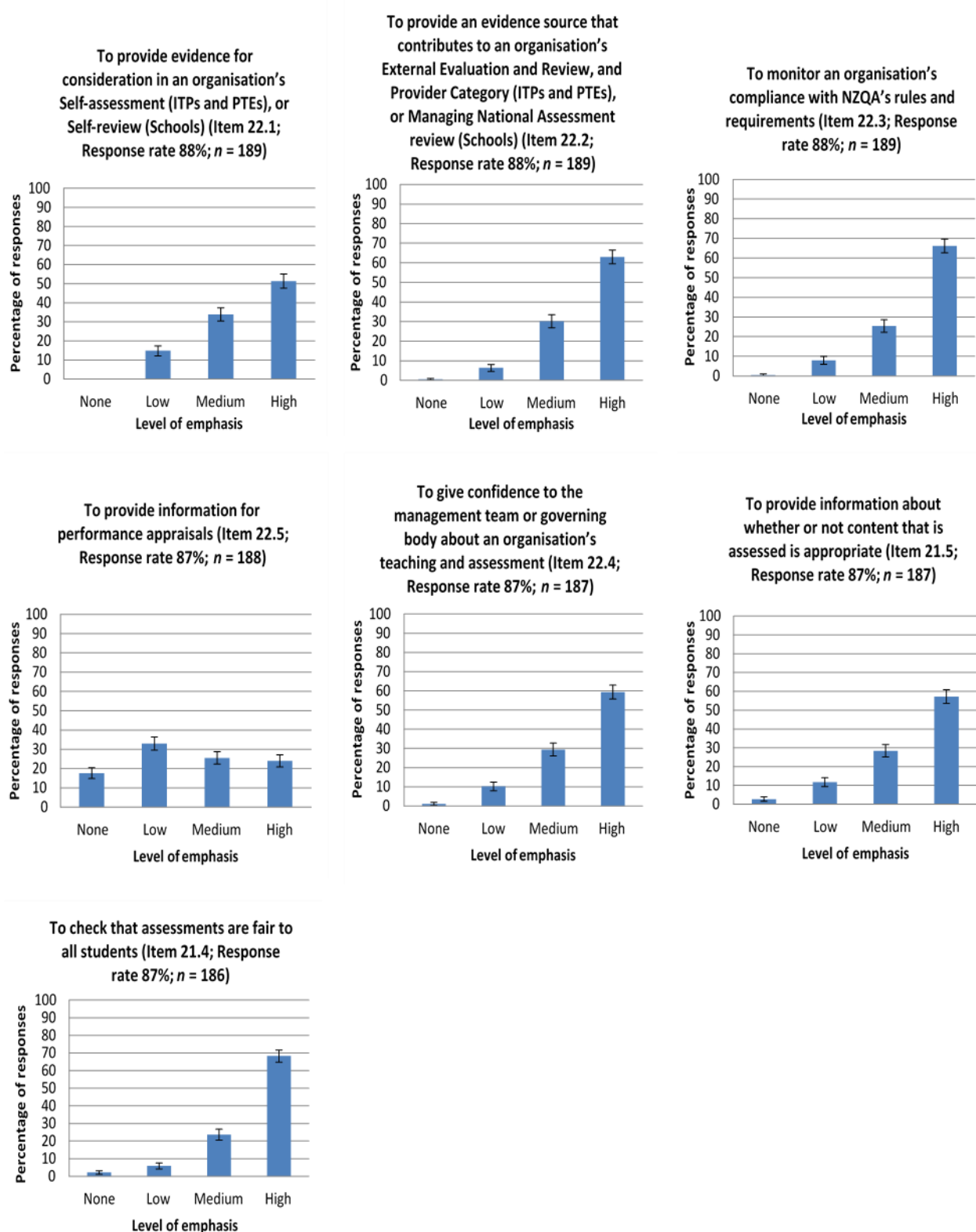
Figure 4.14 The distribution of responses for each NZQA moderation item contributing to the *Organisational quality assurance* factor. Error bars denote standard error of the percentage of responses.

Table 4.15 shows the pattern matrix of NZQA moderation items associated with the *Providing opportunities to learn from NZQA moderation reports* factor. In the exploratory factor analysis, this factor had an initial eigenvalue of 8.9, and accounted for 29.0% of the total variance in the data. Three survey items were associated with this factor, with loadings of varying magnitudes. The weak loading item was also shown to be associated with the factor, *Checking internal assessment quality*.

The concept that unites these three items is the provision of opportunities to learn from NZQA moderation reports. Two items address the learning opportunities provided by moderation feedback and results. The third (26.3: *To give feedback about how well the assessment materials assess achievement against a standard*) reflects the disparity that existed in NZQA moderation reports in terms of providing learning opportunities about assessment materials leading up to and during the data collection for this study (refer 1.3.1). At that time there was variation regarding whether reports provided moderation feedback or results about the assessment materials in a submission, or only about assessor judgements. This lack of uniformity in NZQA moderation reports meant that, depending on the standards moderated, some organisations will have had the opportunity to learn about how well the assessment materials assessed achievement against a standard from some moderation reports, and not from others, and other organisations may not have had such an opportunity at all, depending on the assessment materials they used. Thus, item 26.3 coheres with the unifying concept central to this factor.

Table 4.15   Pattern matrix of NZQA moderation items associated with the *Providing opportunities to learn from NZQA moderation reports* factor.

		Percentage of variance (Extraction Sums of Squared Loadings)	
		28.0	5.2
		Factor	
Items		*Providing opportunities to learn from NZQA moderation reports*	*Checking internal assessment quality*
Q26.2	To provide opportunities for learning from the feedback in moderation reports	.921	
Q26.1	To provide opportunities for learning from the moderation results in moderation reports	.697	
Q26.3	To give feedback about how well the assessment materials assess achievement against a Standard	.377	.368

Note: coefficients less than 0.30 suppressed.

Figure 4.15 shows the distribution of responses to each item associated with the *Providing opportunities to learn from NZQA moderation reports* factor. A clear majority of respondents felt that NZQA moderation emphasised the functions that these items address, with more than 80% rating each item as receiving at least *Medium emphasis*. Of the three items, the level of emphasis on 26.3 (regarding feedback about assessment materials) was perceived most strongly.

To provide opportunities for learning from the feedback in moderation reports
(Item 26.2; Response rate 82%; *n* = 177)

To provide opportunities for learning from the moderation results in moderation reports (Item 26.1; Response rate 82%; *n* = 176)

To give feedback about how well the assessment materials assess achievement against a Standard
(Item 26.3; Response rate 82%; *n* = 177)
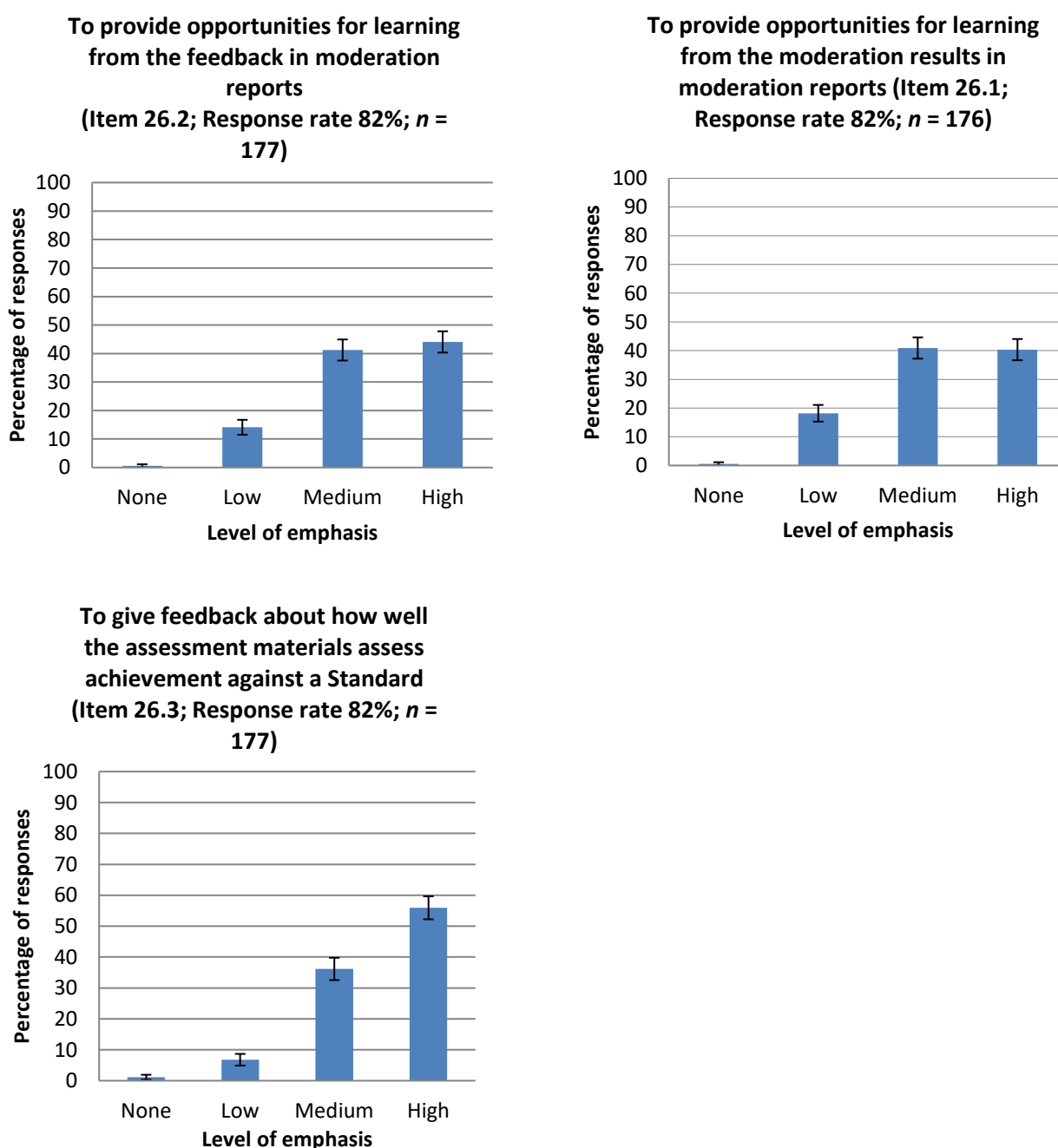
Figure 4.15   The distribution of responses for each NZQA moderation item contributing to the *Providing opportunities to learn from NZQA moderation reports* factor. Error bars denote standard error of the percentage of responses.

4.2.2 Items not loading onto a factor, or omitted from exploratory factor analysis

Two items were not substantially associated with any of the final factors that were extracted in the exploratory factor analyses of NZQA moderation survey items. Item 29.1 (*To inform an organisation's practice*; response rate 81%, *n* = 175) did not load onto a factor in the pattern matrix, although it did load against four different factors in the structure matrix, indicating a correlation with those factors (Tabachnick & Fidell, 2013). Item 29.3 (*To inform reviews of standards*; response rate 80%, *n* = 173) is a conceptual outlier in relation to the rest of the NZQA moderation survey items, as it relates to NZQA's role as a standard-setting body and was the only survey item that did so. For this reason, it was omitted from the exploratory factor analyses.

The question asked for both items regarded its level of importance as a function of NZQA moderation. There was clear endorsement by the survey respondents of the importance of item 29.1, with just over half of the respondents rating it as being of *High importance*, and over one third rating it as being of *Medium* importance. In contrast, the endorsement of item 29.3 was somewhat weaker: Approximately 40% rated it as being of *Medium importance,* 30% as *High importance*, and 25% as *Low importance*.

4.2.3 Measurement scale mean locations of organisation types

Figure 4.16 shows the mean measurement scale locations for the three organisation types on each of the NZQA moderation measurement scales. One scale (*Organisational quality assurance*) stands in contrast to the others, due of the marked difference between the mean location for PTEs compared to those of the other two organisation types. While there appears to be some variation between the locations of the three means on the other scales, none is so dramatic.

PTEs showed more propensity than ITPs or schools to perceive that a greater emphasis is placed on items associated with the *Organisational quality assurance* scale by NZQA moderation, and ITPs showed a similar perception to schools. A one-way analysis of variance reflected these observations, showing significant difference amongst the three means: $F(2, 184) = 11.39$, $p < .001$. Independent samples *t*-tests showed that the difference between the means for PTEs and schools was significant: $t(170) = 4.40$, $p < .001$, as was

the difference between the ITP and PTE means: t (116) = 2.45, p = .02. However, there was no significant difference between the means for ITPs and schools: t (78) = <1, p = .85.[101]

No differences amongst the means for the different organisation types on any of the other four scales were significant. On both the *Providing opportunities to learn from NZQA moderation reports*, and *Improving internal assessment quality* scales, PTEs appeared to have the highest mean location. However, one-way analyses of variance on each scale showed that there were no significant differences amongst the means for the three organisation types. *Providing opportunities to learn from NZQA moderation reports* scale: *F* < 1; *Improving internal assessment quality* scale: F (2, 174) = 1.73, p = .18. Schools had the greatest mean propensity to perceive higher emphasis by NZQA moderation on the items associated with *Maintaining public confidence*, but again no significant differences were shown amongst the means for the organisation types: F (2, 184) = 1.91, p = .15. Lastly, the locations of the means appeared to be similar on the *Checking internal assessment quality* scale; this was borne out with a one-way analysis of variance, which showed no significant difference amongst the three means: F (2, 187) = 1.41, p = .25.

---

[101] The low statistical power of the *t*-tests involving ITPs due to the small ITP sample size is noted.
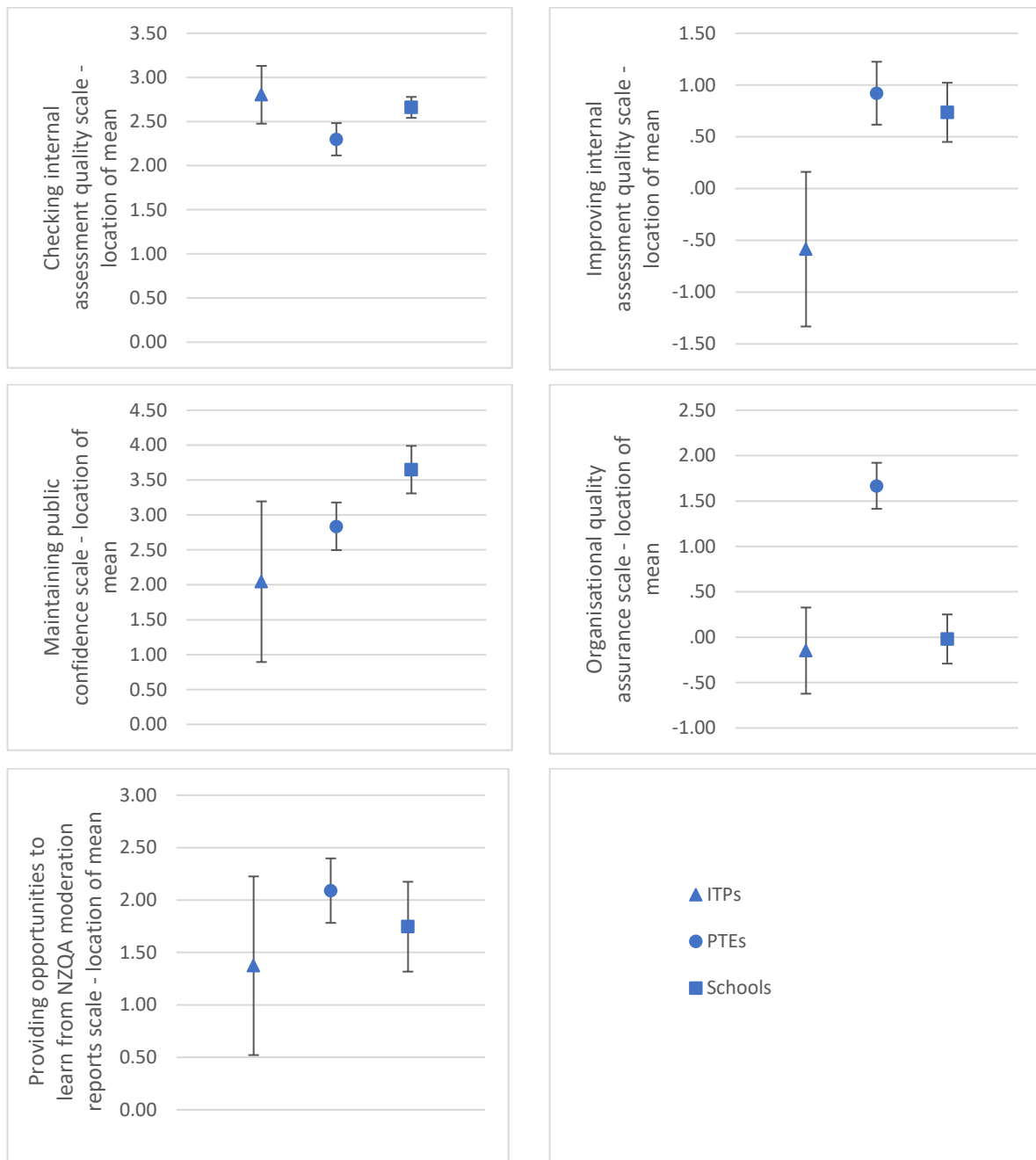
Figure 4.16   Measurement scale location of means for ITP, PTE and school respondents on each NZQA moderation scale. Error bars denote standard errors of the means.

4.2.4 Other tests

4.2.4.1 Pearson's correlations between NZQA moderation scales

Table 4.16 shows the Pearson's correlation coefficients between NZQA moderation scales, for the aggregation of all three organisation types, and for the individual organisation types. There were positive statistically significant correlations of varying strengths between the NZQA moderation scales for the aggregated organisations. While there were significant correlations between all NZQA moderation scales for PTEs, not all correlations for ITPs and schools were significant.

Table 4.16   Pearson's correlation coefficients between NZQA moderation scales for the aggregated organisations, and for the individual organisation types.

| NZQA moderation scale | Organisation type | NZQA moderation scales | | | |
		2	3	4	5
1 Checking internal assessment quality	Aggregate	**.24**	**.29**	**.35**	**.28**
	ITPs		.57	.59	
	PTEs	.26	**.29**	**.41**	**.28**
	Schools			**.46**	**.33**
2 Improving internal assessment quality	Aggregate		**.35**	**.58**	**.70**
	ITPs			.65	**.73**
	PTEs		**.41**	**.62**	**.77**
	Schools		.25	**.54**	**.62**
3 Maintaining public confidence	Aggregate			**.35**	**.30**
	ITPs			.62	
	PTEs			**.45**	**.33**
	Schools			.28	.27
4 Organisational quality assurance	Aggregate				**.43**
	ITPs				.56
	PTEs				**.46**
	Schools				**.39**
5 Providing opportunities to learn from NZQA moderation reports					

Note: Correlations in boldface are significant at the 0.01 level (2-tailed), and correlations in standard font are significant at the 0.05 level (2-tailed). Non-significant correlations have been omitted.

The *Improving internal assessment quality* scale was involved in the two strongest correlations for the aggregated organisations: with the *Providing opportunities to learn from NZQA moderation reports* scale ($r = .70$), and with the *Organisational quality assurance*

scale ($r = .58$). For each of the individual organisation types, the correlations between the scales, *Improving internal assessment quality* and *Providing opportunities to learn from NZQA moderation reports* was significant at the 0.01 level and very strong: for ITPs, $r = .73$, for PTEs, $r = .77$, and for schools, $r = .62$.[102] For the individual organisation types, the correlations between the *Improving internal assessment quality* and *Organisational quality assurance* scales was also strong.

For the aggregated organisations, most correlations involving the *Checking internal assessment quality* scale were weak. Three were the weakest correlations between any NZQA moderation scales: with *Improving internal assessment quality*, *Providing opportunities to learn from NZQA moderation reports*, and *Maintaining public confidence* scales. For the most part, these correlations for the individual organisation types were weak or with lower levels of significance (as per Table 4.16).

There were sizable differences in the strength of correlations between certain NZQA moderation scales for PTEs and schools, with correlations for PTEs being of greater magnitude. The correlations with large differences in magnitude for these organisations were between the *Maintaining public confidence* and *Organisational quality assurance* scales, between the *Maintaining public confidence* and *Improving internal assessment quality* scales, and between the scales, *Providing opportunities to learn from NZQA moderation reports* and *Improving internal assessment quality*.

4.2.4.2 Correlations between NZQA moderation scales and length of service

Pearson's correlation coefficients were calculated between the individual NZQA moderation measurement scales and respondents' length of service either as a teacher, assessor, or administrator in formal education. The only statistically significant correlation found (between the *Organisational quality assurance* scale and respondents' length of service) was weak and negative: $r = -.18$, $p = .02$.

---

[102] The ITP correlations have low statistical power due to the small ITP sample size, and so are not discussed any further.

No significant correlations were found in Pearson's correlation coefficients between the individual NZQA moderation measurement scales and respondents' length of service in academic leadership positions.

4.2.4.3 Chi-square tests of independence on items not accounted for by factors

Chi-square tests of independence conducted on the two items not accounted for by NZQA moderation factors showed that PTE and school respondents rated item 29.1 (*To inform an organisation's practice*) as having slightly more importance than expected under the null hypothesis, but not significantly: $\chi^2(2) = 2.88$. PTE respondents had a greater propensity to rate the item 29.3 (*To inform reviews of standards*) as of higher importance than school respondents, and of higher importance than expected under the null hypothesis: $\chi^2(2) = 6.85$, $p < .05$.

4.2.5 NZQA moderation quantitative results: Summary

In summary, quantitative analyses found that the factors describing how academic leaders saw the functions of NZQA moderation aligned with the four main areas of assessment quality, maintaining public confidence, organisational quality assurance, and professional and organisational learning and development, and illuminated nuance within these areas. Variation was found in the perceived levels of emphasis on items associated with these functions: from almost unanimous recognition of the high emphasis on checking internal assessment quality, to more than half of the respondents rating items concerning increasing the confidence of and developing shared understandings between teachers as receiving at least medium emphasis. Respondents tended to treat the function of checking internal assessment quality as relatively independent or 'stand-alone': For the most part, ratings given to the associated items were only weakly aligned with ratings given to other items. In contrast, a group of other, broader functions (improvement-focused, and with focuses other than assessment) appeared to 'hang together', in that respondents tended to respond to all similarly.

PTE respondents were found to consider the items associated with organisational quality assurance as receiving higher emphasis in NZQA moderation than other respondents did. PTE respondents also tended to be more likely than other respondents to rate the broader

functions (those other than checking assessment quality, i.e., improving assessment, and non-assessment focused functions) in similar ways.

4.3 Quantitative results: Internal moderation and NZQA moderation

Quantitative analyses were conducted comparing internal moderation and NZQA moderation data. Initially, the nature of relationships between the responses regarding the different functions of internal moderation and NZQA moderation were ascertained. Then, where identical items were asked about the two types of moderation, any differences in the importance ascribed to, or emphasis placed on, each were identified.

4.3.1 Pearson's correlations between internal and NZQA moderation scales

Table 4.17 shows the Pearson's correlation coefficients between internal moderation scales and NZQA moderation scales for the aggregated data from the three organisation types, and for the individual organisation types. The correlations varied in strength, and in level of statistical significance.

Table 4.17   Pearson's correlation coefficients between internal and NZQA moderation scales, for aggregated and individual organisation types.

		NZQA Moderation Scales				
		1	**2**	**3**	**4**	**5**
Internal moderation scale	**Organisation type**	Checking internal assessment quality	Improving internal assessment quality	Maintaining public confidence	Organisational quality assurance	Providing opportunities to learn from NZQA moderation reports
3 Improving assessment material and assessor judgement quality	Aggregate	**.33**	.18	**.30**	**.24**	**.25**
	ITPs					
	PTEs	**.40**		**.27**	.20	
	Schools		.25	**.41**	**.32**	**.36**
4 Maintaining public confidence	Aggregate	.15	**.30**	**.44**	**.35**	**.24**
	ITPs			.61		
	PTEs		**.27**	**.43**	**.35**	
	Schools		**.35**	**.37**	**.41**	.26
5 Organisational quality assurance	Aggregate		**.41**	**.34**	**.68**	**.39**
	ITPs		.61		.65	.65
	PTEs		**.48**	**.49**	**.61**	**.38**
	Schools	**.31**	.30	.30	**.69**	**.40**
6 Providing professional learning opportunities	Aggregate		**.31**	**.26**	**.42**	**.26**
	ITPs					
	PTEs		**.35**	**.33**	**.43**	**.35**
	Schools			.30	**.43**	
7 Assisting organisational development	Aggregate		**.36**	**.39**	**.43**	**.30**
	ITPs					
	PTEs		**.34**	**.40**	**.47**	**.29**
	Schools		**.42**	**.47**	**.53**	**.33**
8 Checking educational quality	Aggregate		**.29**		**.57**	**.24**
	ITPs					
	PTEs	**.27**	**.40**	**.33**	**.61**	**.41**
	Schools				**.42**	
9 Improving educational quality	Aggregate	.15	**.34**	**.26**	**.55**	**.35**
	ITPs					
	PTEs		**.36**	**.39**	**.55**	**.37**
	Schools	0.24	.31	.30	**.51**	.29

Note: Correlations in boldface are significant at the 0.01 level (2-tailed), and correlations in standard font are significant at the 0.05 level (2-tailed). Non-significant correlations have been omitted.

The strongest correlation for the aggregated data and the individual organisation types was between the NZQA moderation *Organisational quality assurance* scale and the internal moderation *Organisational quality assurance* scale: for the aggregated organisations, $r =$ .68; for ITPs, $r = $ .65; for PTEs, $r =$ .61; and for schools, $r =$ .69.[103, 104] This was the only correlation between internal moderation and NZQA moderation scales where the coefficient magnitude was similar for all three organisation types.

For the aggregated organisations, PTEs, and schools, the NZQA moderation scale *Organisational quality assurance* also correlated strongly with several other internal moderation scales. It correlated strongly with the internal moderation scale *Checking educational quality* for the aggregated organisations and for PTEs, but moderately for schools.[105] It correlated strongly with the internal moderation scale *Improving educational quality* for the aggregated organisations, PTEs, and schools. For schools, *Organisational quality assurance* also correlated strongly with the internal moderation scale, *Assisting organisational development*. However, the corresponding correlation was moderate for the aggregated organisations and for PTEs.

The scale involved with the weakest and fewest number of statistically significant correlations was the NZQA moderation scale *Checking internal assessment quality*. This was moderately correlated with the internal moderation scale *Improving assessment material and assessor judgement quality*, but only for the aggregated organisations and PTEs. For schools, it was also moderately correlated with the internal moderation scale *Organisational quality assurance*. All other correlations involving this NZQA moderation scale were either weak or non-significant. These findings suggest that the functions encompassed by the NZQA Moderation scale *Checking internal assessment quality* are seen by academic leaders to essentially 'stand-alone' from internal moderation, with little or no relation to it.

---

[103] ITPs and PTEs also each had another correlation of an equivalent strength.
[104] The low statistical power of the ITP correlations is again noted, and ITP correlations are therefore not discussed further.
[105] The difference in magnitude of correlation (of .19) between PTEs ($r =$ .61) and schools ($r =$ .42) is marked.

4.3.2 *z*-Tests for proportions between internal and NZQA moderation items

The results of *z*-tests for proportions on pairs of closed-response items that had internal moderation and NZQA moderation addressing the same function are shown in Table 4.18. For each function, respondents' ratings of internal moderation (in terms of perceived importance) were compared with their ratings of NZQA moderation (in terms of perceived current emphasis). In most cases, internal moderation was shown to be believed to have more importance than NZQA moderation was believed to have emphasis. The only function for which NZQA moderation was seen to hold stronger emphasis (than internal moderation held importance) was maintaining public confidence in national qualifications.

As Table 4.18 shows, for functions relating to assessment quality and assisting organisational development, internal moderation was rated as having both *Medium* or *High importance*, or *High importance*, by significantly greater proportions of respondents than NZQA moderation was rated in terms of the corresponding levels of emphasis: checking that assessments are fair to all students; giving feedback about the quality of assessor judgements; giving feedback about how to improve assessor judgements; assisting in the development of a shared understanding among relevant people; and building the confidence of assessment designers or assessors.

For providing information for performance appraisals, internal moderation was rated as having at least *Medium importance* by a significantly greater proportion than who rated NZQA moderation as having at least *Medium emphasis*. There was no significant difference in the proportions who rated internal moderation or NZQA moderation as having *High importance* (or *emphasis*) for providing information for performance appraisals.

In contrast, for the function of maintaining public confidence in national qualifications, a significantly greater proportion of respondents rated NZQA moderation as having *Medium* or *High emphasis*, or *High emphasis*, than rated internal moderation as having the corresponding levels of importance. However, there was no significant difference in the proportions endorsing internal moderation and NZQA moderation as important or in terms of current emphasis for the function of maintaining public confidence in the internal assessment conducted in an organisation.

Table 4.18 *z*-Test comparisons of endorsement of identical items pertaining to both internal moderation and NZQA moderation

Function	Rating[a]	Internal moderation		NZQA moderation		*z*-score	*p*-value
		Item	Proportion	Item	Proportion		
To check that assessments are fair to all students	High	6.1	0.93	21.4	0.68	6.37	<0.001
	Med or High	6.1	0.98	21.4	0.92	2.63	0.008
To give feedback about the quality of assessor judgements	High	13.3	0.83	26.5	0.53	6.27	<0.001
	Med or High	13.3	0.98	26.5	0.91	3.08	0.002
To give feedback about how to improve assessor judgements	High	13.4	0.78	26.6	0.39	7.66	<0.001
	Med or High	13.4	0.97	26.6	0.75	6.08	<0.001
To assist in the development of a shared understanding among relevant people	High	14.2	0.64	27.3	0.31	6.37	<0.001
	Med or High	14.2	0.94	27.3	0.64	7.43	<0.001
To build the confidence of assessment designers or assessors	High	14.3	0.52	27.4	0.26	5.06	<0.001
	Med or High	14.3	0.89	27.4	0.63	5.88	<0.001
To provide information for performance appraisals	High	7.6	0.22	22.5	0.24	0.53	0.59
	Med or High	7.6	0.61	22.5	0.49	2.48	0.013
To maintain public confidence in the internal assessment conducted in an organisation	High	11.2	0.54	24.2	0.61	1.49	0.14
	Med or High	11.2	0.89	24.2	0.93	1.25	0.21
To maintain public confidence in national qualifications	High	11.3	0.53	24.3	0.7	3.43	<0.001
	Med or High	11.3	0.85	24.3	0.94	2.69	0.007

Note. [a] Internal moderation items were rated in terms of importance; NZQA moderation items were rated in terms of current emphasis

4.3.3 Internal moderation and NZQA moderation quantitative results: Summary

To summarise, academic leaders tended to treat the NZQA moderation function of organisational quality assurance similarly to internal moderation functions with improvement and non-assessment foci (organisational quality assurance, educational quality, and assisting organisational development)—that is, those functions tended to 'hang together' in terms of perceptions of their importance or current emphasis. In contrast, academic leaders appeared to see the NZQA moderation function of checking the quality of internal assessment as independent of internal moderation functions.

Where questions were asked about identical functions of internal moderation and NZQA moderation, internal moderation was rated as having higher importance than NZQA moderation had emphasis for functions relating to assessment quality, assisting organisational development, and providing information for performance appraisals. However, NZQA moderation was rated as having higher emphasis than internal moderation had importance for maintaining public confidence in national qualifications.

4.4 Qualitative results

The qualitative survey data were initially used to assist in finding out what academic leaders perceived the *in-practice* functions of internal and NZQA moderation to be (see 2.6.1). However, during inductive analysis of these data, it became apparent that academic leaders' perceptions of the functions of moderation-in-general (i.e., the *in-principle* functions of moderation) were also revealed (refer 3.4.5.1).

The qualitative results are presented by topic, according to the five main areas in which academic leaders saw moderation to function. The areas are presented in order from those receiving the greatest volume of data to those receiving the least: assessment quality, maintaining public and stakeholder confidence, quality assurance, professional and organisational learning and development, and educational quality. Qualitative results relating to moderation-in-general, internal moderation, and NZQA moderation, are presented for each area. Discussion is also included for each of how those results may relate to the literature, the present study's context, and applicable quantitative findings. For clarity, discussion relating to assessment quality is interspersed with reporting of the qualitative

results. The discussion relating to the other four areas in which moderation is believed to function is presented after the qualitative results for each area. Dissatisfaction with NZQA was also evident in the qualitative survey data, and results evincing this are presented last.

4.4.1 Assessment quality

Assessment quality featured strongly in the functions of moderation according to respondents; it was reportedly addressed by moderation-in-general, internal moderation, and NZQA moderation (see Table 4.19). In the conceptualisation presented in 1.2.5 of the present study occurring in layers of embedded contexts, assessment quality pertains to the context of students' education and learning experiences. The qualitative data showed three dimensions to responses about assessment quality: the component of the assessment process that moderation addresses, the action that moderation takes on that assessment component, and the property of assessment[106] that moderation is concerned with. These dimensions featured in the qualitative data in relation to moderation-in-general, internal moderation, and NZQA moderation. While all three dimensions were addressed through individual quantitative survey items, only two of the dimensions featured in the factors extracted from the quantitative data: the assessment component that moderation addresses (in relation to internal moderation), and the action that moderation takes on the assessment component (in relation to both internal moderation and NZQA moderation).

Table 4.19   Moderation has a role in assessment quality

	Moderation-in-general		Internal moderation		NZQA moderation	
	n	*%*	*n*	*%*	*n*	*%*
Total	**174**		**28**		**10**	
ITPs	*12*	7	*2*	7	*2*	20
PTEs	*100*	57	*18*	64	*7*	70
Schools	*62*	36	*8*	29	*1*	10

---

[106] For explanations of the terminology that will be used for the properties of sound and trustworthy assessment, refer 2.1.3 and 3.4.5.5.

4.4.1.1 Moderation-in-general has a role in assessment quality

A focus on assessment quality was nominated by a large majority of respondents (174) as a main function of moderation (84% of qualitative sample; Table 4.19). The composition of this cohort proportionally mirrors the composition of the qualitative survey sample relatively closely in terms of respondents from each organisation type (refer Table 3.1). The data showed three dimensions to respondents' comments about the role of moderation in assessment quality: the components of the assessment process that moderation focuses on, the actions that moderation has on those assessment components, and the properties of assessment that moderation is concerned with. Each of these dimensions is explicated below.

4.4.1.1.1 Component of the assessment process that moderation-in-general addresses

Assessment-in-general (i.e., assessment practices and processes), assessor judgements, and assessment materials, were all referred to as the component of the assessment process that moderation addresses by participants in their top-of-mind responses about the main functions of moderation (Table 4.20). Of the 174 respondents who included assessment quality amongst the main functions of moderation, 125 referred to moderation addressing assessment-in-general, assessment practices, and/or assessment processes. For example, respondents referred to *"assessment practice"*, *"assessment conduct", "an assessment event"*, or simply to *"assessments"*. Ninety respondents specified that moderation focuses on assessor judgements; referring, for example, to *"[teachers'] assessment decisions"*, *"assessor judgements"*, *"marking"*, *"grading"*, and *"results"*. That moderation addresses assessment materials was specified by 58 respondents. For example, respondents referred to *"assessment activities"*, *"assessment tasks"*, *"assessment tools"*, *"materials used for assessment"*, and *"the assessments designed"*.

Table 4.20   Component of the assessment process that moderation-in-general addresses.

	Assessment-in-general		Assessment materials		Assessor judgements	
	n	%	*n*	%	*n*	%
Total	**125**		**58**		**90**	
ITPs	*11*	9	*2*	3	*1*	1
PTEs	*74*	59	*38*	66	*51*	57
Schools	*40*	32	*18*	31	*38*	42

The strongest focus in the responses (Table 4.20) was on assessment-in-general,[107] which may be indicative that more respondents think of assessment as a whole than as a series of separate component parts in relation to moderation. Alternatively, this strength of focus may be a symptom of using a written survey as the data collection instrument: People tend towards brevity in their responses and are unlikely to specify all components in detail (Cohen et al., 2011). That more respondents indicated that moderation addresses assessor judgements than assessment materials may be due to organisations sometimes using assessment materials that have been supplied to them (discussed further below; refer also 1.2.3). It may also reflect that there generally are greater numbers of moderation decisions about assessor judgements than about assessment materials: For every assessment event there are as many assessor judgements as there are students who were assessed (and possibly with more than one assessor involved), whereas organisations tend to use a limited number of different sets of assessment materials.

Table 4.20 shows that, proportionally, the cohorts referring to moderation addressing the different components of assessment resembled the composition of the qualitative survey sample (Table 3.1) relatively closely in terms of organisation types, although there was some variation in the proportions nominating each component. Respondents from ITPs tended to focus slightly more on assessment-in-general than the other components as separate concerns, possibly suggesting a tendency to think of assessment as a whole rather than as comprising separate parts. Assessment materials appeared to be slightly more of a top-of-mind focus of moderation for PTE respondents than for those in schools, which may reflect the different realities for these two organisation types: Schools can freely access approved assessment materials for a lot of internal assessment via TKI and NZQA (refer 1.2.3), whereas PTEs tend to either develop their own or to purchase materials. Thus, that

---

[107] The more frequent referral to assessment-in-general than to other assessment components is evident throughout the results pertaining to moderation-in-general (item 4).

moderation is a key quality assurance tool for assessment materials is likely to be in more acute focus for PTEs than for schools. Disproportionately more schools were focused on assessor judgements in their top-of-mind thoughts about moderation functions (Table 4.20), for example, *"To provide a forum for feedback and subsequent review of in-school judgements" (School),* which may reflect the pressure that schools are under with NZQA's accountability focus (e.g., national agreement rates; refer 2.4.3.2.1).

4.4.1.1.2 Action of moderation-in-general on assessment quality

Respondents referred to moderation-in-general having a variety of actions on the quality of assessment, as indicated by the verbs they used to describe its functions. These actions included 'assuring', 'evaluating', and 'giving feedback about and improving' (Table 4.21). There was also a small group (17 respondents: 12 PTEs, five schools) who did not specify the action(s) undertaken by moderation (e.g., *"Standardisation of assessment; Fairness in assessment" [PTE]*), meaning that it was not possible to infer the actions that they saw moderation as taking on assessment.

That moderation has a role **assuring** assessment quality was specified by 142 respondents (68% of the qualitative sample), far more than specified it having evaluation or provision of feedback and improvement roles. Of the different parts of the assessment process that moderation could assure, assessment-in-general was nominated by 100 respondents: *"To make sure that the assessment carried out is fair, valid, and consistent, and meets the required standard" (PTE).* Assessor judgements were nominated by 59 respondents: *"Ensuring marking meets national standards … Ensuring consistency of marking and that marking meets standards set" (School),* and 38 respondents nominated assessment materials: *"Ensuring assessment tasks assess the learning outcomes of a unit standard, prescription or similar" (PTE).*

That moderation has an **evaluation** role was specified by 37 respondents, most of whom focused on the evaluation of assessor judgements (Table 4.21). For example, *"To determine that teachers are correctly and consistently identifying the standard and appropriate grade boundaries. Benchmark work nationally" (School).* A similar and smaller number of respondents referred to moderation evaluating assessment-in-general (*"Quality check on assessment components" [PTE]*) and assessment materials (*"To assess appropriateness of [assessment] materials given to students" [PTE]*).

Moderation's role in **providing feedback about and improving** aspects of the assessment process was referred to a small cohort (23; Table 4.21). Of these, more than half focused on assessment-in-general: *"To provide feedback to teachers so that their assessment practices will improve until they meet the competency expected by NZQA" (School)*, and smaller groups focused on assessor judgements and assessment materials.

Table 4.21   Action of moderation-in-general on assessment quality.

		Assure		Evaluate		Provide feedback about and improve	
		n	*%*	*n*	*%*	*n*	*%*
	Total	*142*		*37*		*23*	
	ITPs	*12*	8	*1*	3	*0*	0
	PTEs	*81*	57	*22*	59	*7*	30
	Schools	*49*	35	*14*	38	*17*	74
	Assessment-in-general	*100*		*15*		*14*	
	ITPs	*11*	11	*1*	7	*0*	0
	PTEs	*58*	58	*9*	60	*5*	36
	Schools	*31*	31	*5*	33	*9*	64
	Assessment materials	*38*		*16*		*6*	
	ITPs	*2*	5	*0*	0	*0*	0
	PTEs	*27*	71	*11*	69	*2*	33
	Schools	*9*	24	*5*	31	*4*	67
	Assessor judgements	*59*		*23*		*9*	
	ITPs	*1*	2	*0*	0	*0*	0
	PTEs	*35*	59	*13*	57	*3*	33
	Schools	*23*	39	*10*	43	*6*	67

(Left margin label: Assessment component that moderation addresses)

Note: The totals do not equal the sum of the assessment components as some respondents included more than one component in their response.

Assurance was the action that was most commonly identified in respondents' top-of-mind answers about the functions of moderation (Table 4.21). As an action, 'assuring' is broader than 'evaluating' or 'improving' and may encompass both (refer 2.4.3 and 2.4.4). For example, to ensure that marking meets a certain standard involves checking whether it does, and if it is found not to, making appropriate amendments. In contrast, the actions of 'evaluating', and 'providing feedback about, and improving' are specific: one is about

checking or quality control, and the other is about improvement. Thus, that the most commonly identified action was *assurance* may indicate that most respondents think that moderation functions in this broader way, encompassing both checking and improvement. Alternatively, it may be another example of the brevity that characterises survey responses.

More respondents overall identified *evaluation* than *the provision of feedback and improvement* as an action that moderation takes (Table 4.21), suggesting that more respondents had quality control than improvement as a top-of-mind function of moderation. The quantitative results reflect this: There was almost unanimous endorsement of the importance of the checking assessment quality functions of internal and NZQA moderation, while there was more variation in the level of importance ascribed to their roles in *improving* the quality of internal assessment (see 4.1.1 and 4.2.1).

As shown in Table 4.21, the overall proportions of respondents who included the moderation actions of *assurance* and *evaluation* in their top-of-mind answer about the main functions of moderation resemble the composition of the qualitative survey sample in terms of organisation types (Table 3.1). However, the group who referred to moderation *providing feedback about and improving* aspects of the assessment process contained a disproportionately high number of school and low number of PTE or ITP respondents. This suggests that the feedback and improvement function of moderation features more strongly in school respondents' top-of-mind thoughts about moderation than it does for respondents from the other organisation types. Other differences in focus by organisation type as suggested by the table (e.g., a proportionally greater focus by PTE respondents on *assuring assessment materials*) did not feature in the quantitative data. No significant difference was found in the importance ascribed by respondents from the different organisation types to functions of internal moderation of improving assessment materials and assessor judgements (refer 4.1.4.1) or the perceived emphasis placed by NZQA moderation on functions of checking or improving the quality of internal assessment[108] (see 4.2.4.1). Further, the function of internal moderation checking assessment quality was almost unanimously endorsed (refer 4.1.1), and there was near consensus that NZQA moderation places high emphasis on checking assessment quality (see 4.2.1).

---

[108] The NZQA moderation functions of *Checking internal assessment quality* and *Improving internal assessment quality* both incorporated aspects relating to assessment materials and assessor judgements. Refer 4.2.1.

4.4.1.1.3 Properties of assessment that moderation-in-general is concerned with

Overall, seven properties of assessment were identified by respondents in their top-of-head responses as being considered by moderation-in-general. These were *consistency*, that assessments *assessed what they were meant to assess*, *fairness*, the *robustness of quality*, *accuracy of assessor judgements*, *validity*, and *reliability*.[109] (Table 4.22.) The selection of properties identified aligns with those that Newton and Shaw (2014) classified as comprising the technical quality of assessment (see 2.1.3). The selection is comprehensive and covers those properties that are widely considered in the literature to be important.

Table 4.22   Property of assessment that moderation-in-general is concerned with.

		Consistent		Assess what they are meant to		Fair		Valid		Robust quality		Assr jdgmts are accurate		Reliable	
		n	%	n	%	n	%	n	%	n	%	n	%	n	%
	Total	*113*		*88*		*59*		*55*		*42*		*40*		*13*	
	ITPs	*9*	8	*2*	2	*8*	14	*10*	18	*5*	12	*0*	0	*4*	31
	PTEs	*72*	64	*56*	64	*44*	75	*35*	64	*32*	76	*21*	53	*9*	69
	Schools	*32*	28	*30*	34	*7*	12	*10*	18	*5*	12	*19*	48	*0*	0
Assessment component addressed	**Assmt-in-general**	*73*		*55*		*35*		*34*		*32*		*0*		*8*	
	ITPs	*9*	12	*1*	2	*6*	17	*8*	24	*5*	16			*2*	25
	PTEs	*47*	64	*36*	65	*26*	74	*23*	68	*23*	72			*6*	75
	Schools	*17*	23	*18*	33	*3*	9	*3*	9	*4*	13			*0*	0
	Assmt mats	*8*		*39*		*11*		*14*		*11*		*0*		*3*	
	ITPs	*0*	0	*1*	3	*2*	18	*2*	14	*0*	0			*1*	33
	PTEs	*8*	100	*26*	67	*8*	73	*9*	64	*10*	91			*2*	67
	Schools	*0*	0	*12*	31	*1*	9	*3*	21	*1*	9			*0*	0
	Assr jdgmts	*48*		*0*		*19*		*11*		*0*		*40*		*3*	
	ITPs	*0*	0			*1*	5	*0*	0			*0*	0	*1*	33
	PTEs	*30*	63			*14*	74	*6*	55			*21*	53	*2*	67
	Schools	*18*	38			*4*	21	*5*	45			*19*	48	*0*	0

Note: Assmt-in-general = Assessment-in-general; Assmt mats = Assessment materials; Assr jdgmts = Assessor judgements. Total sample for each category shown in boldface. The totals do not equal the sum of the properties, as some respondents included more than one property in their response.

---

[109] For the explanation of why results for validity and reliability are reported here, refer to 3.4.5.5.

Consistency was the most commonly identified assessment property that moderation addresses, with 113 respondents including it in their top-of-head responses (Table 4.22). Most (73 respondents) referred to consistency in relation to assessment-in-general: *"Ensure consistent assessment practices" (PTE)*. Consistency was referred to in relation to assessor judgements by 48 respondents: *"Checking marking is consistent across subjects, the school and nationally" (School)*. A small number (eight PTEs) referred to it in relation to assessment materials. Consistency over time, between assessors, between organisations, and nationally featured in responses, (e.g., *"To ensure that assessment decisions are … consistent, in terms of the learning outcomes and standards set (within campuses, across campuses and nationally)" [PTE])*.

Many of the aspects that respondents nominated for consistency (e.g., of assessor judgements between assessors, organisation, nationally, and over time) are recognised in the literature as being key facets of reliability of internal standards-based assessment (refer 2.1.3). However, some respondents appeared to see consistency more broadly, applying it to aspects not necessarily thought of as being associated directly with reliability of assessment *per se* (e.g., assessment practices or assessment materials), although they may assist with ensuring reliability.[110] As per Table 4.22 (compared to the qualitative sample composition; Table 3.1), PTE respondents appear to have been more likely than school respondents (in particular) to think of consistency as having application to functions other than to assessor judgements (i.e., and relate to assessment-in-general or assessment materials). It is noted that the one quantitative survey item that specifically addressed assessment consistency (internal moderation item 6.5: *To check that assessor judgements about assessment evidence are consistent*) was almost unanimously endorsed as being of high importance (refer 4.1.1), indicating that any differences in top-of-mind response were not reflected in the importance ascribed to the internal moderation function concerning the consistency of assessor judgements.

That **assessments assess the requirements that they were meant to**, and that these assessments are at the correct level, was specified by 88 respondents as being an assessment property of concern to moderation. As per Table 4.22, this property was referred to in relation to assessment-in-general (55; e.g., *"To ensure that teaching and assessing meet the assessment criteria; To ensure that learner work meets the NZQA standards" [PTE])*, and assessment materials (39; e.g., *"Assessment (e.g., the task and materials) is*

---

[110] Refer to 2.2.3.1 for discussion of the challenges involved with reaching reliable assessor judgements in internal assessment, particularly when portfolio-approaches are used.

consistent with required requirements of standard (and curriculum Achievement objectives)"
[School]). However, it was not specifically referenced in relation to assessor judgements.

Respondents identified the requirements to be assessed as those of unit standards, courses, or other prescriptions. Some respondents also made specific mention that the assessment was set at the required level: *"Ensuring assessment tasks assess the learning outcomes of a unit standard, prescription or similar; Checking assessment tasks cover all learning outcomes; Checking assessment tasks are at the correct level of the unit, prescription or similar" (PTE).*

The property of assessments 'assessing what they are meant to' and at the correct level coheres with the literature in terms of commonly accepted properties comprising sound and trustworthy assessment (refer 2.1.3). The identification of the types of standards, courses, and prescriptions that define the relevant requirements for assessment reflect those currently in use in the New Zealand secondary and tertiary contexts (refer 1.2.1). The cohort of respondents for which this property was top-of-mind relatively closely resembled the composition of the qualitative survey sample in terms of organisation types (Table 3.1). These qualitative top-of-mind results were reflected in the quantitative results: The items that address aspects of this property (internal moderation items 6.2, 6.3, 13.1, 13.2: refer 4.1.1; NZQA moderation items 21.1, 21.2, 26.4, 27.1: refer 4.2.1), and the relevant scales (internal moderation: *Improving assessment materials and assessor judgements*, refer 4.1.3; NZQA moderation: *Checking internal assessment quality, Improving internal assessment quality*, refer 4.2.3) showed no significant difference in the ratings of importance or emphasis given by respondents from the different organisation types.

Moderation considers how **fair** assessment is, according to 59 respondents. Thirty-five respondents referred to fairness in relation to assessment-in-general: *"Fair assessment" (PTE).* Fairness of assessor judgements was identified by 19: *"Ensure marking of students' work is fair and appropriate in terms of level etc" (PTE).* Eleven referred to fairness in relation to assessment materials:

> *Moderation checks to ensure assessments are fair, valid and consistent, and the language used in asking questions is clear and at the appropriate level for the subject in question, [and will] not unnecessarily disadvantage any one student due to lack of clarity in what is required of them. (PTE)*

As per Table 4.22, it appears that fairness was not top-of-mind for many school respondents when they thought of moderation: Only 12% of the respondents who identified fairness were

from schools, far fewer than would be required to resemble the composition of the qualitative sample (Table 3.1). The reasons for this are unclear. Perhaps other processes are used in schools to ensure that assessments are fair; or perhaps there is a lack of clarity about the meaning of 'fairness' and ways in which moderation can address it. This disparity in top-of-mind response was reflected in the quantitative results pertaining to NZQA moderation, but not those pertaining to internal moderation. In the NZQA moderation results, the relevant item (21.4, *To check that assessments are fair to all students*) was associated with *Organisational quality assurance* (refer 4.2.1). Schools were found to consider the items associated with *Organisational quality assurance* to receive less emphasis by NZQA moderation than PTEs did (see 4.2.3). In contrast, the internal moderation item (6.1) that pertained to checking that assessments are fair to all students was almost unanimously endorsed as having high importance (refer 4.1.1).

Validity of assessment was identified by 55 respondents as a property that is considered by moderation. Thirty-four respondents explicitly used the terms 'valid' or 'validity' in their top-of-mind answer in relation to assessment-in-general (e.g., *"To ensure fairness, <u>validity</u>, and consistency in assessments" [ITP;* emphasis added]). Fourteen did so in relation to assessment materials (e.g., *"Ensures the assessments designed are … valid (academically sound and they assess what they say they assess and nothing else) …" [PTE]*), and 11 did so in relation to assessor judgements (e.g., *"Ensure assessment judgements are valid" [School]*).

That validity was explicitly identified by this cohort of respondents may reflect the ubiquitous nature of the term in the context of assessment and moderation. However, in most cases it was not possible to ascertain exactly what each respondent meant by 'validity' when they used the term (see 3.4.5.5), and it is unlikely that all respondents understood it to mean the same thing. Thus, it is not possible to infer the meaning of these responses, although this is an example of the challenges posed to clear communication of not having one commonly accepted, universal definition of validity (refer 2.1.3).

That the assessment is of **robust quality** was another assessment property identified as being of concern to moderation in 42 respondents' top-of-head answers (Table 4.22). Respondents included this in relation to assessment-in-general (32) and assessment materials (11), but not in relation to assessor judgements specifically. Their comments encompassed aspects such as robustness: *"To ensure assessment is robust … "* (ITP); quality: *"To maintain a high standard of assessment and delivery within the education sector"* (PTE); credibility: *"To apply standards to ensure that the assessment process is credible"*

(PTE) and *"Ensures the assessments designed are fit for purpose (meet industry requirements)…" (PTE)*; integrity: *"To maintain the integrity of assessment" (School)*; and relevance: *"To ensure that an activity being used is appropriate for the specific context and group of students that it is being used with" (School).*

The property of assessments being of robust quality was not top-of-mind for many school respondents, particularly as compared to respondents from PTEs: A disproportionately low number of school respondents and a disproportionately high number of PTE respondents included it in their initial answers (see Table 4.22, as compared with Table 3.1). Perhaps other processes in schools are used to address the quality of assessment, where PTEs may rely on moderation to fulfil this function. Alternatively, properties other than the quality of assessment might be what most school respondents think of first when they consider moderation. The quantitative results did not reflect the apparent discrepancies between organisation types in the level of top-of-mind focus on robust quality: The one directly relevant item (internal moderation item 8.1: *To check that the assessment practices in an organisation are robust*) was endorsed by almost all respondents as having high importance (refer 4.1.2).

The **accuracy of assessor judgements** was nominated by 40 respondents as being a concern of moderation (Table 4.22). This group comprised an almost equal number of schools compared to PTEs, but no ITPs. For example, *"[moderation is concerned with the] standard and correctness of marking and grading" (PTE)*, and *"[moderation functions] to ensure accuracy and consistency in marking of NCEA" (School).*

In contrast to the other identified assessment properties, the accuracy of assessor judgements was top-of-mind for more school respondents and fewer ITP and PTE respondents than would proportionally resemble the qualitative sample (Table 3.1). This may reflect differences in the secondary and tertiary contexts, such as the levels of public and political interest in the moderator-assessor agreement rates (refer 2.4.3.2.1). However, any apparent differences in top-of-mind awareness of assessor judgement accuracy did not translate into the quantitative results. No real differences between the organisation types were evident in terms of how important internal moderation was seen to be regarding the accuracy of assessor judgements, or how much emphasis NZQA moderation was seen to place on this. Internal moderation item 6.4 (*To check that assessor judgements about assessment evidence are accurate*) was almost unanimously rated as having high importance (refer 4.1.1). There were no significant differences in the emphasis perceived by respondents from the different organisation types on the two NZQA moderation functions of

Checking and *Improving internal assessment quality*, both of which accuracy of assessor judgement items were associated with (see 4.2.1 and 4.2.3).

Reliability was identified as an assessment property of concern to moderation by 13 respondents in their top-of-mind answers. These respondents explicitly used the terms 'reliable' or 'reliability' in relation to assessment-in-general, assessment materials and assessor judgements: *"[Moderation functions to] ensure rigour, <u>reliability</u>, and consistency of assessments" (ITP;* emphasis added).

As per Table 4.22, reliability featured in the top-of-mind answers for a small number of ITP and PTE respondents, but none from schools. Little can be read into this discrepancy, because the properties of consistency and accuracy of assessor judgements in internal assessment encompass most (if not all) of what reliability means (refer 2.1.3). However, this finding does suggest that the term 'reliability' may be in more common usage in the tertiary sector than in the secondary sector.

4.4.1.2 Internal moderation has a role in assessment quality

In the second section of the survey, respondents were given the opportunity to comment on the functions of internal moderation, in practice. Twenty-eight respondents (13% of qualitative sample; see Table 4.19) referred to internal moderation having a role regarding assessment quality in their comments. While two PTE respondents expressed reservations, the rest implicitly or explicitly endorsed this function. For example, internal moderation, *"[functions to] ensure externally purchased assessment material meets required standard" (ITP);* and *"[checks] that the needs of varying student demographics are being met by the assessment process, e.g., English as a second language or students who have difficulty with reading/writing" (PTE).* Further, *"[internal moderation] processes often involve staff assessing [student evidence] and then another staff member verifying their judgements, or a staff member writing a new assessment task and then another staff member checking that it complies with the requirements of the standard" (School).*

The three dimensions that featured in the qualitative data relating to moderation-in-general and assessment quality also featured here: the component of the assessment process that moderation addresses, the action that moderation takes on that assessment component, and the property of assessment with which moderation is concerned. Respondents specified that internal moderation addressed different components of assessment: assessment-in-

general (15), assessment materials (13), and assessor judgements (six). Respondents variously referred to internal moderation having different actions (i.e., assuring, evaluating, improving) on the components of assessment. Eight respondents (one ITP, three PTEs, four schools) referred to internal moderation *assuring* assessment quality: *"It's the only way in our present system to ensure validity" (School)*. Seven respondents (five PTEs, two schools) referred to it *evaluating* assessment quality: *"[Internal moderation] assists in determining whether the assessment meets the necessary criteria for the awarding of the grade" (PTE)*. Five respondents (one ITP, four PTEs) referred to it *improving* assessment quality: *"Identifying improvements in assessment activities" (PTE)*. Ten respondents (eight PTEs, two schools) did not specify the action that internal moderation took regarding assessment quality (e.g., *"Consistency amongst organisations" [PTE]*).

As reported in 4.1.1, analysis of quantitative data resulted in three factors (*Checking assessor judgement quality*, *Improving assessment material and assessor judgement quality*, and *Checking assessment material quality*) being extracted as functions of internal moderation, and the importance of the items associated with these factors was very strongly endorsed. However, these factors featured weakly in the qualitative data regarding internal moderation. Only five respondents (four PTEs, one school) commented on internal moderation having a role in evaluating (checking) the quality of assessment material: *"[internal moderation] assists in determining whether assessment measures what it intends to measure; assists with determining whether assessment meets learning outcomes; …" (PTE)*. Only three (one PTE, two schools) commented on it having a role in evaluating (checking) the quality of assessor judgements: *"Moderating is supposed to check the assessor's ability to assess at the national standard" (School)*. Only four (one ITP, three PTEs) commented on it having a role in improving the quality of assessment materials and assessor judgements: *"The moderator's comments should be considered to improve all assessments" (ITP)*. Perhaps the weak presence of these factors in the qualitative data indicates that most respondents felt that the quantitative items had covered these functions adequately, and therefore did not feel the need to emphasise, or supplement through comment the ratings they had given the items.

Respondents referred to four properties of assessment that internal moderation addressed: that the assessment *assesses what it is meant to assess*, is of *robust quality*, is *consistent*, and is *fair*. All four properties had also been identified as being of concern for moderation-in-general.

Whether an assessment **assesses what it was meant to assess** and at the required level was commented on by ten respondents (one ITP, six PTEs, three schools) as a concern of internal moderation. For example: *"Internal moderation provides the provider with a check on the quality of the assessment materials prepared in terms of meeting the requirements of the course/standards" (PTE)*.

That assessment was of **robust quality** was commented on by nine respondents (seven PTEs, two schools) as being of concern to internal moderation. These respondents determined that internal moderation considered aspects such as the educational soundness: *"[Among other things, it checks] that the program and assessment instruments are educationally sound" (PTE)*; quality: *"Current practice of internal and external moderation is focused strongly on the standard of assessments" (PTE)*; and relevance: *"[The assessment] should meet and reflect the special character and relevance to the school's community" (School)*; of assessments, as well as the facilitation of assessments: *"Internal moderation is a vital part of the structure that ensures efficient and robust facilitation of assessment" (School)*.

Consistency of assessment was noted as being a concern of internal moderation by seven respondents (three PTEs, four schools). Consistency of marking, comparability of grades awarded, and consistency between organisations featured in the data: *"Internal moderation should provide confidence that the assessment results are comparable to similar assessment results in any other organisation" (School).* However, one respondent felt that internal moderation was ineffective in maintaining national consistency: *"… the results of internal moderation cannot be relied on for national consistency…" (PTE).* As with moderation-in-general (see 4.4.1.1.3), most of these aspects align with key facets of reliability, although several respondents appeared to apply the property of consistency more broadly than what is commonly accepted as reliability (e.g., consistency of teaching between organisations).

Fairness also featured in the qualitative data as an assessment property of concern to internal moderation, with five respondents (four PTEs, one school) mentioning it: *"As such, by using the internal moderation process, we can ensure that our results are robust, <u>fair</u> and consistent" (PTE;* emphasis added).

All four assessment properties that respondents commented on as being of concern for internal moderation were addressed to some extent by one or more closed-response survey items. This may account for the low number of respondents addressing the properties in

their comments; perhaps most respondents did not feel the need to emphasise or elaborate on them. Other explanations might include that respondents had answered the initial survey item asking for their top-of-mind thoughts about the functions of moderation so fulsomely that they did not feel the need to repeat themselves, or alternatively, that the sparsity of response is an example of the brevity that characterises survey responses.

4.4.1.3 NZQA moderation has a role in assessment quality

Only ten respondents commented on the function of NZQA moderation regarding assessment quality in the open-field items (see Table 4.19; 5% of qualitative sample), suggesting that most respondents did not have anything further to add to the content addressed by the closed-response items. Eight (two ITPs, six PTEs) endorsed that NZQA moderation did have a function regarding assessment quality (although some qualified their endorsement), while two (one PTE, one school) asserted that NZQA moderation was ineffective in this role.

Seven of the respondents (one ITP, five PTEs, one school) referred to NZQA moderation focusing on assessment-in-general (including assessment practices and methodology), while two specifically mentioned assessor judgements as its focus. No respondent made mention of assessment materials in relation to NZQA moderation, contrasting with comments about internal moderation (4.4.1.2) and moderation-in-general (4.4.1.1). This may reflect the recent emphasis on assessor judgements (and not assessment materials) by much of NZQA moderation (refer 1.3.1). In the quantitative analysis, items addressing assessment materials and those addressing assessor judgements were associated with each of the two factors relating to assessment quality that were extracted (*Improving internal assessment quality* and *Checking internal assessment quality*, refer 4.2.1), suggesting that respondents did not consider materials and judgements as separate in relation to NZQA moderation, where for internal moderation they did.

Respondents attributed different actions to NZQA moderation in terms of its effect on assessment quality (as was the case for moderation-in-general—4.4.1.1, and internal moderation—4.4.1.2): assuring (four), evaluating (none), and improving (four). Three did not indicate the actions taken. Of those who commented on NZQA moderation having a role in improving the quality of internal assessment, one ITP begrudgingly agreed: *"Institutions do not like receiving reports that require action plans to be submitted and approved by NZQA; however, this does require that professional conversations are held to facilitate the required*

changes" (ITP). The other three (two PTEs, one school) disagreed that NZQA moderation fulfils an improvement role in relation to internal assessment quality: *"There is very little emphasis [in NZQA moderation] on helping schools improve their performance" (School).* Thus, of the two NZQA moderation factors extracted (*Improving internal assessment quality* and *Checking internal assessment quality*, refer 4.2.1), only the *improvement* factor featured in the NZQA moderation qualitative data. Respondents had perceived there to be greater emphasis on the quantitative items associated with *Checking internal assessment quality* than on the items associated with *Improving internal assessment quality* (see 4.2.1). The difference in perceived emphasis suggests that respondents generally accepted NZQA moderation functioning to *check* or *evaluate* assessment quality (and did not feel the need to comment further regarding this) but held more diverse views regarding it functioning to *improve* assessment quality and were prompted to reply to the closed-response items that suggested it did.

Respondents wrote of NZQA moderation being concerned with different assessment properties, as they had for moderation-in-general (4.4.1.1) and internal moderation (4.4.1.2). The properties of *consistency*, *robustness of quality*, *fairness*, and that assessments *assess what they were meant to*, were referenced.

Consistency was nominated by five respondents (one ITP, four PTEs) as being a concern to NZQA moderation. This included consistency in marking and in assessment methodology, both within and between organisations. For example, *"to ensure fairness and consistency in assessment methodology used by different providers across the country—particularly where the providers are offering the same course" (PTE).*

Assessment having **robust quality** was referred to by five respondents (four PTEs, one school). Three of the PTEs felt that NZQA moderation was concerned with this: *"[NZQA] moderation should ensure assessment is robust" (PTE).* However, the other two respondents were dubious that NZQA moderation was concerned with this property: *"Some of the responses [in NZQA moderation reports] have more to do with minor details than the overall educational value of the assessment" (PTE).*

Fairness was suggested as a concern for NZQA moderation by two PTE respondents: *"To ensure that we grade our students' work in a fair and consistent manner" (PTE).*

Whether internal assessments **assess what they are meant to assess** as a concern of NZQA moderation was suggested by one PTE respondent: *"…NZQA moderation in our*

experience is there primarily for standard adherence…" (PTE). The property of assessments assessing what they are meant to concerns standard adherence, which is a key role of NZQA moderation. Perhaps most respondents felt that this property was adequately addressed by the quantitative survey items, and, therefore, did not feel the need to reiterate it. A vast majority of respondents believed that NZQA moderation strongly emphasises the relevant quantitative items (21.1 and 21.2; refer 4.2.1), supporting this inference.

Another critical function of NZQA moderation concerns the *accuracy of assessor judgements*: a property not raised by respondents in relation to NZQA moderation but referenced in relation to moderation-in-general (4.4.1.1) and internal moderation (4.4.1.2). It may be that respondents felt that this function was adequately addressed in the closed-response items (specifically item 21.3; see 4.2.1), and so chose not to comment on it. Item 21.3 (*To check that assessor judgements are accurate against the requirements of the standard*) was almost unanimously viewed by respondents as receiving high emphasis, which supports this inference. Perhaps the lack of comments was also a symptom of 'response fatigue'—of respondents getting tired of completing the survey, and so tending towards brevity in their open-field answers.

4.4.2 Maintaining public and stakeholder confidence

Maintaining public and stakeholder confidence was prevalent in respondents' comments about the functions of internal moderation and NZQA moderation, but only featured weakly in their top-of-mind thoughts about the functions of moderation-in-general (see Table 4.23). More respondents commented on internal moderation and NZQA moderation having a role in maintaining public and stakeholder confidence than did on any other function, which suggests that respondents felt strongly about this role. Sentiments ranged from endorsement to dissent.

Along with the public, stakeholders whose confidence was to be maintained (as specified by respondents) included current and prospective students, parents and families, communities, employers, industry, regulators including NZQA, and the government. In the conceptualisation presented in 1.2.5 that positions the present study in layers of embedded contexts, these stakeholders inhabit the overarching societal context layer. Respondents variously saw internal and NZQA moderation as functioning to maintain public and stakeholder confidence in a range of aspects, including internal assessment conducted by an organisation, an organisation's quality assurance processes, the quality of education

provided by an organisation, the qualifications awarded by an organisation, an organisation itself, the ability of graduates, the education system, and (mentioned only in relation to NZQA moderation), the National Qualifications Framework.

Table 4.23   Moderation has a role in maintaining public and stakeholder confidence

	Moderation-in-general		Internal moderation		NZQA moderation	
	n	%	n	%	n	%
Total	*10*		*54*		*24*	
ITPs	*2*	20	*4*	7	*1*	4
PTEs	*2*	20	*39*	72	*20*	83
Schools	*6*	60	*11*	20	*3*	13

4.4.2.1 Internal moderation has a role in maintaining public and stakeholder confidence

The maintenance of public and stakeholder confidence featured strongly in the data from the internal moderation section of the survey, with 54 respondents (Table 4.23; 26% of qualitative sample) commenting on it.

The notion of internal moderation functioning to maintain public and stakeholder confidence was broadly supported by 36 respondents. Of these, 26 (20 PTEs, six schools) endorsed the role without proviso. For example, *"[a role of internal moderation is to] provide confidence to external stakeholders, e.g. [to] assure industry that graduates are reaching the standards of achievement that education providers are claiming" (PTE),* and *"we want NZQA to have confidence in us as a private training provider" (PTE).*

Ten other respondents (three ITPs, five PTEs, two schools) articulated that internal moderation had more of an indirect role, via being part of a quality assurance system. For example,

> *I believe that members of the public as a rule are not particularly interested in the details such as moderation. … However, they do want to know that processes and outcomes are rigorous, and that education institutions are held in good esteem. They also want to know that processes are fair. (ITP)*

On the other hand, 16 respondents disagreed that internal moderation had a role in maintaining public or stakeholder confidence. Of these, 13 respondents (one ITP, ten PTEs, two schools) explicitly dissented. For example, *"I don't believe internal moderation achieves this as the differences in assessment practices and the assessments themselves between institutions are too great" (School).* Six (five PTEs, one school) asserted that the maintenance of public and stakeholder confidence was more of the role of national (e.g., NZQA) moderation: *"Internal moderation sometimes does not have access to the external national standard. [National] moderation is needed to provide that function" (PTE).* This assertion was reflected by only some of the quantitative data. NZQA moderation was seen to place more emphasis on maintaining public confidence in national qualifications than was perceived important in internal moderation, but not on maintaining public confidence in the internal assessment conducted in organisations (refer 4.3.2).

Sixteen respondents (three ITPs, 12 PTEs, one school) suggested that internal moderation is not publicly visible (in terms of the process or results), or that the public and stakeholders have little (or no) knowledge of it. For example, *"I doubt that the general public would have any idea about the importance of internal moderation and/or its relevance to quality or otherwise of education and qualifications" (ITP).*

In contrast, three respondents (one PTE, two schools) explicitly described how their organisations made the moderation process or results public, with the inference that their endorsement of internal moderation having a role in maintaining such confidence was influenced by this. As one respondent explained, *"The students understand that [internal moderation] happens and that they may have to wait for assessments to be returned because internal moderation has to occur; but they accept this as necessary and it helps them trust the system" (School).*

4.4.2.2 NZQA moderation has a role in maintaining public and stakeholder confidence

Twenty-four respondents (refer Table 4.23; 12% of qualitative sample) commented on NZQA moderation having a role in maintaining public and stakeholder confidence.[111] That NZQA moderation has a role in maintaining public and stakeholder confidence was endorsed by 11

---

[111] In addition to these 24 respondents, five others asserted that national moderation had more of a role in maintaining public and stakeholder confidence than internal moderation did (see 4.4.2.1).

respondents (one ITP, eight PTEs, two schools). For example, *"If assessment is not robust, we are doing all stakeholders an injustice. [NZQA moderation] should ensure assessment is robust, and if it is the system will produce fit-for-purpose graduates who will instil confidence in the system" (PTE).*

Nine respondents (eight PTEs, one school) expressed dissatisfaction about NZQA moderation having this role or qualified their support of it. These respondents variously felt that there were shortcomings in the NZQA moderation process, that it was ineffective in some way, or that it led to organisational responses that adversely impact on educational quality. For example,

> *NZQA place a lot of emphasis on the 'perception' of assured quality but appeared to be hampered by their systems. For example, I would be calling for 'specific' reported assessment scripts for moderation rather than any three assessing a unit. (PTE)*

Dissent from the notion that NZQA moderation functions to maintain public or stakeholder confidence was expressed by five PTE respondents. For example, *"The public are more interested in the knowledge received, and as the NZQA moderation is rarely consistent, it is an unreliable benchmark" (PTE).* Five respondents (one ITP, three PTEs—including one that had dissented with NZQA moderation having this role, one school) opined that NZQA moderation is not publicly visible, or the public know little (if anything) about it: *"Moderation is a 'back end' process not generally seen by public, parents or employers" (PTE).*

4.4.2.3 Moderation-in-general has a role in maintaining public and stakeholder confidence

In contrast to the number of respondents commenting about this function in relation to internal and NZQA moderation (4.4.2.1, 4.4.2.2), when asked at the start of the survey to list the main functions of moderation, only ten (5% of qualitative sample; Table 4.23) identified maintaining public or stakeholder confidence. For example, *"it also allows parents, students and employers to have confidence in NZ qualifications" (School),* and *"[moderation functions] to assure NZQA that the institution is managing national assessment" (School).*

4.4.2.4 Discussion

As per Table 4.23, markedly more respondents commented about moderation maintaining public and stakeholder confidence after they had been asked specifically about internal or NZQA moderation doing so (via closed-response items in the internal moderation and NZQA moderation survey sections), than mentioned it beforehand. This suggests that maintaining public and stakeholder confidence was not a function that sprang to mind immediately for most respondents. Moderation acts directly on internal assessment, and the maintenance of public or stakeholder confidence is a function that is somewhat removed from the actual act of moderation; it occurs in a context that is not immediate to that within which moderation occurs (see 1.2.5). It seems likely that such indirect functions of moderation are not what most respondents initially think of when they consider the role of moderation-in-general. Note that no inference regarding the importance of (or emphasis on) this function can be drawn from the finding that few respondents identified it in their top-of-mind responses. As analysis of the quantitative data found, respondents clearly endorsed the importance of all internal moderation items addressing the maintenance of public confidence and considered NZQA moderation to emphasise these functions, with at least 85% of respondents rating each item as having at least *Medium importance* or *emphasis* (refer 4.1.1 and 4.2.1).

The range of views expressed about internal or NZQA moderation having a role in maintaining public and stakeholder confidence indicated that there were diverse and sometimes contradictory perspectives held among respondents. The volume of respondents that referred to either type having such a role indicates that those respondents felt strongly enough (for whatever reason) to comment. This apparent strength of feeling may be due, in part, to the importance of reputation and public and stakeholder confidence for organisations, and the reality that a change in either can have financial and other impacts for an organisation. The publication of NZQA reviews of organisations (EERs for ITPs and PTEs and MNAs for schools, refer 2.3.1.2, 2.3.1.3.1) may have heightened respondents' sensitivities to the maintenance of public and stakeholder confidence. The sensitivity of school respondents may have been heightened further due to the annual publication of schools' NCEA results in the mainstream media (e.g., via Stuff's "School report," n.d.), and publication of the annual National Agreement Rate (see 2.4.3.2.1). The level of feeling about NZQA moderation having such a role may also be indicative of a sense of frustration with, or the perception of a lack of fairness about, the situation particularly if respondents had misgivings about the NZQA moderation process or requirements. Although proportionally more PTE and fewer school respondents commented (as compared to the composition of the qualitative sample; Table 3.1), this apparent discrepancy between organisation types

was not reflected in the quantitative data. The quantitative analyses revealed that there were no significant differences between academic leaders from the three organisation types in terms of the importance placed on the items associated with *Maintaining public confidence* by internal moderation (refer 4.1.3), or in the perceived emphasis placed on these items by NZQA moderation (see 4.2.3).

4.4.3 Quality assurance

Quality assurance (beyond that of internal assessment) was another area in which moderation was seen to function. The quality assurance role of moderation-in-general was perceived variously as focusing on the organisation (particularly in relation to externally-set requirements), the education system, and on qualifications.[112] Interpreted through the embedded layers of context conceptualisation (1.2.5), these apply to the education organisation and education system contexts. Internal and NZQA moderation were both considered to have roles in organisational quality assurance, and as such, were seen to focus at the organisation contextual layer. (Refer Table 4.24.)

Table 4.24   Moderation has a role in organisational and wider quality assurance

	Moderation-in-general		Internal moderation		NZQA moderation	
	n	%	*n*	%	*n*	%
Total	*44*		*27*		*7*	
ITPs	*1*	2	*3*	11	*1*	14
PTEs	*21*	48	*18*	67	*6*	86
Schools	*22*	50	*6*	22	*0*	0

4.4.3.1 Moderation-in-general has a role in quality assurance

Quality assurance was included in the top-of-mind response of 44 respondents when asked to identify the main functions of moderation (see Table 4.24; 21% of qualitative sample). Some respondents referred to quality assurance, control, or improvement, without further elaboration, while others referred to quality assurance of the organisation (in relation to externally set requirements), of qualifications, or of education system-level aspects.

---

[112] For results pertaining to moderation's function in the quality assurance of internal assessment, refer 4.4.1.

Quality assurance, control, or improvement was identified as one of the main functions of moderation by 21 respondents, with further no explanation given as to what the quality assurance, control, or improvement related to or focused on. Within this group, quality assurance was referenced by 12 (seven PTEs, five schools; e.g., *"Quality assurance"* [School], *"To ensure quality"* [PTE], and *"Moderation is a quality management process"* [PTE]). Quality control was nominated by nine (two PTEs, seven schools; e.g., *"Quality check"* [School], *"Quality control check"* [School], and *"Monitoring of standards"* [School]), and one respondent referred to both quality control and improvement: *"Maintain [and] improve quality" (PTE)*.

For 11 respondents (one ITP, six PTEs, four schools), the quality assurance focus of moderation-in-general was at the organisation level. These respondents variously discussed the organisation's overall quality (e.g., *"To ensure [the] school is at the national level"* [School]), the organisation's self-assessment (e.g., *"Self-assessment of current practice"* [PTE]), providing confidence to internal stakeholders (e.g., *"Provide transparency and confidence to teachers, … school managers etcetera"* [School]), quality improvement (e.g., *"It is also a chance to help identify areas of improvement within systems, materials and tutors"* [PTE]), and its internal quality assurance requirements (*"Quality assurance of assessment activities ... in relation to guidelines and rules laid down by the school ..."* [School]).

Externally set requirements relating to quality assurance were referred to by 10 respondents (eight PTEs, two schools). Some observed that moderation is a requirement imposed by NZQA, (e.g., *"[Moderation is] part of [the] NZQA registration and accreditation requirements"* [PTE]). Others explicitly suggested that a reason their organisation engaged with moderation was to satisfy that requirement (e.g., *"Ensure compliance with NZQA, TEC and governmental regulations and requirements"* [PTE]). One respondent asserted that internal moderation functioned to assist an organisation to meet national moderation requirements: *"Internal moderation ensures that we perform well during [national] moderation" (PTE)*. Another alluded to the measures that organisations will take to ensure compliance: *"I am currently under pressure to inspect the scripts [samples of assessed work] before sending them out [for NZQA moderation] to avoid the embarrassment of failing [national] moderation" (PTE)*.

Quality assurance of qualifications was a main function of moderation according to nine respondents (three PTEs, six schools). For example, *"To maintain the integrity of National Qualifications (School)*.

Quality assurance that focused at the education system-level featured in nine respondents' answers (two PTEs, seven schools). Four (two PTEs, two schools) referred to comparability across the sector. For example, *"The moderation process ensures the integrity of the qualification is maintained and that all educational institutions have appropriate and equivalent standards allowing for consistency in the education sector" (PTE).* Two schools saw moderation as supporting the calculation of a national agreement rate between moderators and assessors (e.g., *"To provide a vehicle that allows for the national agreement rates to be derived" [School]*), and a further two expressed cynicism about the system-level quality assurance function of moderation: *"I think it is to ensure comparability and to assist schools in knowing they are on track, BUT I fear it is more to do with number counting by [the] NZQA and government" [School;* emphasis in original].

4.4.3.2 Internal moderation has a role in organisational quality assurance

The role of internal moderation in organisational quality assurance was commented on by 27 respondents in the internal moderation section of the survey (see Table 4.24; 13% of qualitative sample). In their comments, respondents variously focused on internal moderation being part of an organisation's quality management system, being used to inform other organisational processes, being an externally set requirement, having a role in confirming quality, and that organisations can address its findings.

Eleven respondents (two ITPs, five PTEs, four schools) referred to internal moderation as being a part of quality management, (e.g., *"[Internal moderation] is part of a [quality assurance] system" [PTE]*). Three noted that internal moderation can be used to check other internal processes, (e.g., *"We also use [internal moderation] to quality check our internal [Quality Management System] systems" [PTE]*).

Nine respondents (one ITP, eight PTEs) noted that informing other organisational processes is a use to which internal moderation findings can be put. For example, improving internal processes (*"I believe internal moderation is an opportunity for an organisation to improve their systems and practices" [PTE]*), contributing to an organisation's self-assessment, and informing human resources processes (*"We find internal moderation provides valuable information contributing towards staff performance appraisals" [PTE]*). Two respondents explicitly linked internal moderation to national moderation: One appeared to use the findings to inform the selection of the sample of assessed work to be sent for external moderation (*"Internal moderation is also used as a source of information for national moderation" [PTE]*),

and the other felt that internal moderation had a positive impact on their organisation's national moderation results (*"… we know [our national moderation results] are made better by the internal systems"* [PTE]).

Small numbers of respondents observed that engaging in internal moderation was an externally-set requirement which was imposed on organisations (three), considered that a function of internal moderation was to confirm the quality and performance of their organisation (two), or referred to the opportunity that organisations have to address internal moderation findings in a timely manner (two).

4.4.3.3 NZQA moderation has a role in organisational quality assurance

Seven respondents referred to NZQA moderation having a role in an organisation's quality assurance in their qualitative data (refer Table 4.24; 3% of qualitative sample).

Compliance with NZQA requirements was referred to by four respondents (one ITP, three PTEs). Respondents noted that compliance was the primary focus for the organisation (*"NZQA moderation is generally seen as a compliance requirement with the reports containing minimal feedback to assist the development required"* [ITP]), as well as for NZQA (*"NZQA [Tertiary Assessment and Moderation][113] main focus (sic) is on compliance. The other functions appear to be supportive of this primary goal"* [PTE]). Another respondent expressed an opposing view, asserting that, while compliance was important, it should not drive an organisation's own quality assurance processes; instead: *"… I believe it is the organisation's responsibility to take the lead in its own assurances (moderation) of these matters"* (PTE).

Concern about the fallibility of the NZQA moderation system was expressed by two PTE respondents, who also expressed concern about the focus of NZQA's moderation and how resource-intensive it is for organisations to comply with. For example, *"While its intent is to strengthen confidence, it actually results in negative criticism about [organisations] wasting resources dotting miniscule i's and crossing irrelevant t's—using cumbersome processes that slow the nation's productivity instead of enhancing efficiency"* (PTE).

---

[113] See 1.3.1 for information about the Tertiary Assessment and Moderation business unit.

An organisation's self-assessment (refer 2.3.1.3.1) was mentioned by two other PTE respondents. They observed that NZQA moderation provides an 'external view' which, presumably, can be utilised as an evidence source for the organisation's own self-assessment. One PTE respondent expressed concern at the ramifications of non-compliance with NZQA's external evaluative review requirements, and the ongoing nature of those ramifications.

4.4.3.4 Discussion

Quality assurance was included by one fifth of respondents in their top-of-mind answers regarding the main functions of moderation, the second most common function identified. This prominence in respondents' minds was reflected in the quantitative data: *Organisational quality assurance* was a factor that was extracted both from the internal moderation data (see 4.1.1) and from the NZQA moderation data (refer 4.2.1). Strong endorsement or recognition was given to almost all items associated with these two factors: Over 90% of respondents rated each of the internal moderation items as being of at least *Medium importance*, and 85% or more rated each of the NZQA moderation items as receiving at least *Medium emphasis*. Further, respondents tended to rate the items associated with both factors in similar ways: Once calibrated onto measurement scales, internal moderation *Organisational quality assurance* and NZQA moderation *Organisational quality assurance* were found to be strongly and positively correlated (see 4.3.1, Table 4.17).

In contrast to the strong support for most items associated with the two *Organisational quality assurance* factors, there was far weaker endorsement of, or perceived emphasis placed on, the one item for each type of moderation that addressed the provision of information for performance appraisals (7.6 for internal moderation, and 22.5 for NZQA moderation).[114] It is noted that in the qualitative data the only explicit reference to performance appraisals or management was made in relation to internal moderation.

The accountability purpose of quality assurance seems to have featured relatively strongly in respondents' top-of-mind responses about the functions of moderation. The responses relating to quality control and compliance with internal and externally set requirements, are clearly and primarily accountability-focused (refer 2.4.3), whereby only a small number of

---

[114] Although there was stronger support for the importance of internal moderation providing information for performance appraisals than there was perceived emphasis on this by NZQA moderation (refer 4.3.2).

responses clearly showed quality improvement as the prime purpose. This may be reflective of Law's (2010; 2.4.4) assertion, and Ehren and Hatch's (2013; 2.4.5) finding, that quality improvement was a secondary focus of quality assurance, whereas accountability purposes are usually the primary focus. However, responses relating to overall quality of the organisation, internal stakeholder confidence, system-level quality assurance, and quality assurance of qualifications (which comprise most of the data)[115] can be seen to incorporate both accountability and improvement purposes (refer 2.4.5). These data suggest that some academic leaders do not subscribe to an either/or view of the purposes that moderation serves, and instead hold a more encompassing or holistic view.

The data pertaining to internal moderation having an organisational quality assurance function seem to suggest more of a quality improvement-focus than the moderation-in-general data, although they are not clear-cut in terms of accountability versus quality improvement purposes. Where some data associated with organisational quality assurance (e.g., the responses referring to internal moderation being used to check internal systems) appear to indicate a quality control purpose, it is not possible to infer the purpose ascribed to other data associated with this function. Similarly, where some data associated with the function of informing other organisational processes (e.g., the responses referring to using moderation findings to inform internal process improvements) reflect a quality enhancement purpose, other data associated with this function (e.g., responses referring to using moderation results in performance appraisals) are more aligned with an accountability purpose. These data lend weight to the inference that 'accountability versus improvement' as a purpose is somewhat of a false dichotomy (see 2.4.5).

The element of the qualitative sample who referenced NZQA moderation having an organisational quality assurance function was small and the data suggest that this function is primarily perceived to serve accountability. Most of the data (those focusing on compliance with external requirements, and expressing concern) align with accountability, and its closely related purposes of compliance and quality control, reflecting what literature acknowledges as primary functions of external quality assurance within the education sector: accountability and quality control (refer 2.4.3).

---

[115] Twelve respondents (i.e., approximately one quarter of the cohort who identified quality assurance in their top-of-mind functions of moderation) identified quality assurance as a function without further elaboration. Without such explanation, it is not possible to determine the purpose(s) that these respondents ascribed to quality assurance (refer 2.4.3, 2.4.4, and 2.4.5).

A disproportionately large number of school respondents, and disproportionately few from ITPs or PTEs, included quality assurance as a function of moderation-in-general (refer Table 4.24), as compared to the qualitative survey sample composition (see Table 3.1). The reasons for these differences are unclear, although the differences in publicly-reported accountability measures that organisations in each sector are exposed to, and the different levels of media, political, and public interest in each, may contribute (e.g., national agreement rates, 2.4.3.2.1, and mainstream media reporting of NCEA pass rates for schools via Stuff's "School report," n.d.; versus NZQA confidence status and EER report for individual TEOs, 2.3.1.3.1). Furthermore, schools' internal assessments tend to contribute primarily to the (large scale) NCEA qualifications, where those of TEOs contribute to a range of qualifications and limited-credit courses, which may explain the apparent higher level of focus of school respondents on the quality assurance of qualifications and system-level aspects (as compared with that of respondents from the tertiary sector). However, this apparently greater awareness among school respondents of the role of moderation-in-general in quality assurance was not reflected in the quantitative data regarding the importance of either internal or NZQA moderation having this function. For both types of moderation, PTE respondents tended to rate the items associated with *Organisational quality assurance* as having significantly higher importance or emphasis than ITP or school respondents did (see 4.1.3 and 4.2.3).

Within the samples commenting that internal or NZQA moderation has a role in organisational quality assurance, PTE respondents emphasised using internal moderation to inform other organisational processes. It is noted that no school respondents made mention of doing so, although school respondents tended to affirm internal moderation as being a part of an organisation's quality assurance system. These differences may suggest that internal moderation results are used in a more wide-ranging way in PTEs than they are in schools and ITPs. Perhaps in schools and ITPs, internal moderation is seen as one component within the organisation's quality assurance system, with the primary focus of assuring the quality of internal assessment, whereas in PTEs it is seen as contributing more broadly to the organisation's quality. Further, perhaps in PTEs more so that in the other organisation-types, articulation is seen between internal moderation and other organisational quality assurance processes, with results from one process informing others as appropriate. No school respondents commented on NZQA moderation having a role in organisational quality assurance, which could suggest that this function does not feature strongly for them, or that they felt that the coverage of the closed-response items was adequate. It could also suggest that NZQA moderation only plays a component part with a relatively narrow focus (on the quality of internal assessment) in schools' organisational quality assurance systems.

4.4.4 Professional and organisational learning and development

Professional and organisational learning and development was another broad area that was seen to be addressed by moderation, and which relates to the organisation context and embedded contexts within that (1.2.5). Respondents affirmed that internal moderation has a role in providing opportunities for professional learning, organisational development, and supporting communities of practice. Moderation-in-general was seen to provide opportunities for professional learning and to support communities of practice. However, while the quantitative data showed that respondents considered NZQA moderation to provide professional learning opportunities, the qualitative data mainly showed respondents rejecting this notion. The qualitative data also showed some respondents commenting on other professional learning and development opportunities provided by NZQA (beyond those provided via moderation itself). See Table 4.25.

Table 4.25   Moderation has a role in providing professional and organisational learning and development opportunities

	Moderation-in-general		Internal moderation		NZQA moderation		NZQA: other opportunities	
	n	%	*n*	%	*n*	%	*n*	%
Total	*24*		*41*		*12*		*10*	
ITPs	*0*	0	*4*	10	*2*	17	*0*	0
PTEs	*15*	63	*26*	63	*8*	67	*6*	60
Schools	*9*	38	*11*	27	*2*	17	*4*	40

4.4.4.1 Internal moderation has a role in professional and organisational learning and development

Professional and organisational learning and development as functions of internal moderation were commented on by 41 respondents (refer Table 4.25; 20% of qualitative sample). Specifically, academic leaders referred to internal moderation offering opportunities for professional learning and development, contributing to communities of practice, and offering opportunities for organisational development. According to these respondents, the learning and development opportunities focused on assessment, moderation, and teaching and student learning.

That internal moderation provides opportunities for professional learning and development was strongly endorsed, with 22 respondents (two ITPs, 13 PTEs, seven schools) commenting as such. For example, it was referred to as offering opportunities for *"professional development of staff involved" (PTE)*, and *"professional/capability development in teaching teams" (ITP)*. Some respondents specifically referred to the opportunities as being for new teachers and assessors (*"Provide professional support for new staff, …"* [School]), where others referred to the opportunities as applying to all staff (*"Training all staff in assessment and moderation…" [PTE]*).

Internal moderation contributing to communities of practice was commented on by 15 respondents (two ITPs, seven PTEs, six schools). The opportunities provided for collaboration, sharing practice, and engaging in professional conversions were specifically mentioned. For example, *"Our internal moderation process is a peer process, which enables cross fertilisation of ideas and practices across the faculty [and] allows for a component of staff development" (PTE);* and *"[internal moderation provides the opportunity for sharing] of good practice between staff within our school teaching the same subject and between staff at different schools when we collaborate—especially in small subject areas" (School).*

Internal moderation providing opportunities for organisational development was mentioned by five respondents (four PTEs, one school). These opportunities included for developing or improving assessment or moderation practices and systems: *"[Internal moderation gives us the opportunity to develop] improved moderation practices" (PTE).*

Four respondents (two ITPs, two PTEs) commented that internal moderation enabled the professional development needs of staff to be identified: *"[Internal moderation] can highlight areas where our assessors may need upskilling or support" (PTE).*

Assessment, moderation, teaching, and student learning were identified as foci of learning and development opportunities offered by internal moderation. Learning opportunities about assessment were specified by 11 respondents (one ITP, eight PTEs, two schools): *"[Internal moderation provides] professional development for tutors to extend their knowledge and understanding of different unit standards and assessment methods" (PTE).* Respondents from four PTEs also commented that the opportunities were about moderation: *"Staff improve their understanding of both assessment and moderation—it constitutes effective professional development" (PTE).* Learning and development opportunities also focused on teaching and student learning, according to seven respondents (one ITP, three PTEs, three schools): *"Internal moderation should also provide the opportunity to review the [student]*

learning and teaching materials as assessment is not just about learners' achievement. It is also about the [student] learning and teaching process" (ITP). However, as one respondent observed, these opportunities for learning and development were sometimes missed:

> *Feedback on the teaching and learning programme can be provided as a consequence of the moderator observing the student work. At times it is very obvious where a teaching point has been missed by the students as they all will be weak in that area. So, feedback around the teaching and learning programme is relevant and can be a significant learning point for the teacher. Moderating is supposed to check the assessor's ability to assess at the national standard; however, as a former moderator, it also provides a significant opportunity for professional development which is currently lost. (School)*

Seven respondents who referred to the learning and development opportunities offered by internal moderation were not positive in their sentiments. Three school respondents qualified their endorsement, pointing out that: internal moderation was not the main avenue through which these opportunities were generated in their school; the opportunities were sometimes lost (refer the quote above); or that the level to which internal moderation afforded these opportunities varied across the school. For example,

> *All of these [internal moderation] processes mean that staff are involved in professional conversations about the standard and the student work, thus providing opportunities for [professional learning and development] around assessment practices, assessment judgments, and content. In some parts of the school this process works really well, [but] in others moderation is seen as a compliance process and less dialogue is entered into between the staff members. (School)*

Two respondents (one PTE, one school) opined that internal moderation did not afford these opportunities, due to workload pressures, moderation focus, or skill deficits. For example, *"Internal moderation is not a collegial activity. In its current form, it is actually a heavy workload that distorts the overall workload of teachers. Collegial conversations and learning do not happen in this rushed and overworked environment" (School).* A further two PTEs expressed non-specific negative sentiment regarding professional learning opportunities afforded by internal moderation.

4.4.4.2 Moderation-in-general provides opportunities for professional and organisational learning and development

The role of moderation in providing opportunities for professional and organisational learning and development was referenced by 24 respondents at the start of the survey (see Table 4.25; 12% of qualitative sample). Specifically, these respondents saw moderation-in-general as providing professional learning and development opportunities, contributing to communities of practice, and supporting organisational learning and development.

Moderation-in-general was seen to provide opportunities for professional development by 15 respondents (eight PTEs, seven schools), including for teachers, tutors, and assessors: *"To develop tutor knowledge of unit standards and requirements. To develop tutor confidence and ability to independently correctly assess and mark learner work" (PTE).* Three explicitly mentioned what the learning and development opportunities were about, identifying assessment, unit standard requirements, and teaching content, while the rest included only that the opportunities were provided, (e.g., *"To provide professional development for teachers" [School]*).

Moderation-in-general was seen to play a role in communities of practice by 12 respondents (eight PTEs, four schools). One respondent described it thus: *"[Moderation is] one of the bridges between teachers (and support staff) who are delivering courses and formulating and marking assessments" (PTE).* These respondents considered moderation to provide a forum for professional conversations (e.g., *"To provide an opportunity for discussions and feedback around best practice and quality management systems" [PTE]*) and sharing ideas: *"The process of teachers to internally sharing idea, their own knowledge and experience, understanding between each other, for consistency improving of academic excellent (sic)" (PTE).* One saw moderation as helping to develop a shared language (*"To instil a language that all team members will understand" [PTE]*). Others stated that moderation helps with a consistent shared understanding and practice: *"To ensure consistent understanding of a standard before it is used to assess learning. … To ensure consistency of understanding and applying an activity across a team of teachers before it is used" (School).*

Six respondents (four PTEs, two Schools) saw moderation-in-general as supporting organisational learning and development, by *"[providing] a resource of benchmarked exemplars" (School)*, strengthening connections and communication between staff, and providing a forum through which to focus on quality assurance processes. One saw the

development opportunities provided to be broad: *"[Moderation offers an] organisational opportunity to strengthen and or celebrate effectiveness of teaching, assessment relativity (sic) to outcomes and student progress"* (PTE).

4.4.4.3 NZQA moderation's role in providing professional learning opportunities

Whether or not professional learning opportunities are afforded by NZQA moderation was commented on by 12 respondents (Table 4.25; 6% of qualitative sample). Ten (two ITPs, six PTEs, two schools) rejected that NZQA moderation provides any such opportunities, primarily because of the nature, and lack, of feedback in the NZQA moderation reports (e.g., *"Only the results of the moderation of materials and learner samples are provided in the report." [PTE]*), a lack of suggestions of ways to improve, (e.g., *"[the moderation] feedback [provided] usually does not give suggestions for fixes!! [ITP]*), and variability in the quality of moderation feedback:

> The quality of the moderator's feedback would need to be more consistent for the above [closed-response items regarding NZQA moderation providing opportunities for learning and development] to be high emphasis. I have previously spoken about how non-collegial this system is. (School)

Some respondents expressed generally negative sentiments regarding NZQA moderation providing any learning opportunities (e.g., *"The learning opportunities are undermined by moderator's reports that are of very little value." [School]*).

In contrast, one PTE respondent acknowledged that there is scope for learning opportunities to be provided by NZQA moderation via the moderation reports, in that *"constructive feedback"* could be given. Another simply asserted that professional learning opportunities were available through NZQA moderation.

4.4.4.4 Other professional learning and development opportunities provided by NZQA

Ten respondents commented on other professional learning and development opportunities provided by NZQA (Table 4.25; 5% of qualitative sample). Eight of the 10 expressed dissatisfaction with these other opportunities.

Six (two PTEs, four schools) referred to the 'best practice' or 'moderation' workshops run by NZQA, although the tenor of their comments differed according to respondents' organisation type. While both PTE respondents acknowledged that NZQA does run workshops for providers (e.g., *"Moderation workshops offer training to support our skills and understanding of units" [PTE]*), those from schools complained about access to the workshops or their cost. For example,

> *NZQA offers best practice workshops in generic subjects in bigger cities where there are the numbers attending. THIS DOES NOT HELP THE SMALLER country schools as they are unable to access subject specific best practice workshops—too costly travel & relief and smaller numbers so courses often cancelled in smaller centres (sic). (School;* capitalisation in original*)*

Four other PTE respondents expressed the view that the support and guidance provided by NZQA was inadequate. For example, *"[NZQA] conducts regular audit and monitor and review and judgement, but totally lack of dis-function of Support for Education Provider (sic)"* (PTE).

4.4.4.5 Discussion

In the qualitative data, the number of respondents who commented on internal moderation providing professional and organisational learning and development opportunities was second only to the number commenting on it having a role in maintaining public confidence. As was so in respect of maintaining public confidence, more respondents commented on internal moderation proving these learning and development opportunities after the closed-response items had explicitly addressed these functions than was so beforehand. Both the number of responses (for internal moderation) and the marked increases in responses after prompting indicate that most respondents felt the need to affirm the role of internal moderation in this. It may not have been immediate for almost half (they only commented once prompted), nonetheless, most endorsed internal moderation in this role when prompted.

Professional and organisational learning and development also featured in the quantitative data as functions of internal moderation, with two factors extracted: *Providing professional learning opportunities*, and *Assisting organisational development* (refer 4.1.1). Quantitative analyses revealed that these two factors, once calibrated onto measurement scales, were strongly and positively correlated: Respondents tended to rate the items associated with

each in a similar way (4.1.4.1). The functions addressed by the items associated with the two internal moderation factors were endorsed by a clear majority of respondents, with approximately 90% rating almost all as having at least *Medium importance* (refer 4.1.1). Quantitative analyses revealed that there was little difference across organisation types in the ratings of importance of items associated with internal moderation *Providing professional learning opportunities* and *Assisting organisational development* (refer 4.1.3).

For 12% of the qualitative sample, providing learning and development opportunities and contributing to communities of practice featured among their top-of-mind functions of moderation, indicating that these functions had an immediacy for those respondents. However, no ITP respondents included them as top-of-mind functions, suggesting that the provision of learning opportunities or supporting communities of practice were not functions that these respondents immediately associated with moderation. The reasons for this are unclear; once prompted through the closed-response items, the four ITP respondents endorsed through their comments that internal moderation does have these functions.

Most findings relating to moderation-in-general and internal moderation providing opportunities for professional and organisational learning and development reflect aspects of the literature about moderation for learning and improvement purposes. These include that the learning and support can apply to all teachers and assessors, that the learning and improvement opportunities can relate to any aspect of the assessment, teaching, and student learning cycle, the learning and improvement potential that resides in collegial conversations, collaboration and sharing practice, and identifying learning needs (see 2.4.4.2). One respondent summed up the professional and organisational learning potential of internal moderation, thus: *"By communicating and reaching consensus as a department through the process of internal moderation, you are creating professional confident teachers. You are creating a healthy department that is not afraid to teach students" (School).* In doing so, this respondent alluded to many of the findings in the literature: the learning that occurs through dialogue and negotiation, the impact that this learning can have on teachers' professional confidence, and the potential for positive flow-on effects into teaching practice and the teaching environment (see 2.4.4.2). While data were not collected about the approaches to internal moderation that each organisation uses, it is evident from respondents' comments that a number use a social moderation approach, at least in part. It is possible that this contributes to the stronger emphasis of internal moderation than of NZQA moderation on improvement and learning.

The qualitative data revealed distinctly stronger endorsement of internal moderation providing professional learning and development opportunities than NZQA moderation. Of respondents who commented on internal or NZQA moderation respectively regarding this, three quarters affirmed that internal moderation did provide these opportunities, where almost all disagreed that NZQA moderation provided such opportunities. The quantitative data for the two functions for which internal and NZQA moderation could be directly compared reflected this difference: More respondents saw internal moderation as having higher importance than the corresponding levels of emphasis by NZQA moderation for assisting in the development of a shared understanding among relevant people, and in building confidence of assessment designers or assessors (see 4.3.2).

The dissent regarding NZQA moderation providing professional learning and development opportunities may be reflective in part of the expert model of moderation that NZQA mainly uses (refer 2.4.2). In expert moderation approaches, many of the learning opportunities arise through moderation feedback, and opportunities for learning are impacted by the clarity, content, and tone of that feedback (refer 2.4.4.2.8). It is noted that dissatisfaction with the moderation feedback provided by NZQA featured strongly in the qualitative data. Furthermore, the focus on assessor judgements and not assessment materials in some NZQA moderation in recent years (refer 1.3.1) may have contributed to the level of negative sentiment expressed. The dissent and dissatisfaction expressed about NZQA moderation reports appears to be in contrast with the findings of the Controller and Auditor-General (2012), and Hipkins (2013), both of whom found that the majority of school teachers considered NZQA moderation feedback to be helpful (see 2.4.4.2.8.1). It is unclear whether this apparent contrast reflects that only those who felt strongly enough to express their dissent commented in the present survey, a change in NZQA's moderation reports, or another reason (e.g., different study samples: The previous studies canvased the opinions of teachers, where the present study canvased academic leaders). The quantitative data also suggest that NZQA moderation provides learning opportunities via moderation reports, with the factor *Providing opportunities to learn from NZQA moderation reports* one of those extracted (refer 4.2.1). However, it is noted that a clear majority of respondents perceived that the three items associated with this factor were emphasised by NZQA moderation, with over 80% rating each as having at least *Medium emphasis* (refer 4.2.1). Quantitative analyses also revealed that there was little difference in the emphasis perceived by respondents from the three organisation types on these items (4.2.3).

The finding that some respondents recognise that NZQA provides professional learning and development opportunities via avenues other that moderation (4.4.4.4) can be seen to reflect

the claims of various scholars that moderation can address different stages of the assessment process, beyond the assessment instrument and assessor judgements (see 2.4.0.1). For example, moderation may address assessment designers' understandings of a standard prior to the development of an assessment instrument, or it may calibrate assessors' judgements prior to marking students' work. Most who commented on these other opportunities expressed dissatisfaction with them: the cost, lack of access, or inadequacy of support provided. These sentiments imply that respondents would like more and better access to professional learning and development opportunities from NZQA. As suggested by studies including Crimmins et al. (2016; refer 2.4.4.2.1), participants tend to recognise and value the professional learning and development opportunities that are offered by moderation, including those which address a range of assessment stages. The general tenor of respondents' comments about NZQA moderation and other professional learning and development opportunities provided by NZQA echo the observations of Hipkins et al. (2016) that there is little in NZQA's approach to moderation to ensure that professional learning opportunities are provided (see 2.4.4). One is left with the impression that, at the time of data collection for the present study, this group of respondents was frustrated by a perception that NZQA was not more effective at providing professional learning opportunities.

4.4.5 Educational quality

Educational quality was another area that respondents saw as being addressed by moderation. As well as elaborating on the role of internal moderation in maintaining educational quality, respondents saw that moderation-in-general had this function (see Table 4.26). Respondents referred to aspects of educational quality relating to programme, curriculum, and teaching, and to aspects relating to student learning and achievement. In contrast, educational quality barely featured in relation to NZQA moderation, with only a few respondents mentioning it. Interpreted in terms of the conceptualisation presented in 1.2.5, this function of moderation concerns the embedded contexts within education organisations.

Table 4.26   Moderation has a role in maintaining educational quality

	Moderation-in-general		Internal moderation		NZQA moderation	
	n	*%*	*n*	*%*	*n*	*%*
Total	*32*		*20*		*6*	
ITPs	*3*	9	*2*	10	*2*	33
PTEs	*25*	78	*13*	65	*3*	50
Schools	*4*	13	*5*	25	*1*	17

4.4.5.1 Moderation-in-general has a role in maintaining educational quality

The maintenance of educational quality was nominated by 32 respondents when they listed the main functions of moderation (see Table 4.26; 15% of qualitative sample). Respondents differentiated between aspects of educational quality pertaining to programme, curriculum, and teaching, and those pertaining to student learning and achievement.

Twenty-seven respondents (one ITP, 22 PTEs, four schools) referenced moderation-in-general addressing the quality of the programme, curriculum, and teaching. These respondents variously focused on courses and programmes, content (*"[moderation functions] as a teaching tool to show what needs to be taught and the range of that teaching" [PTE]*), teaching practice, resources, and activities. Respondents commented on ensuring that these met the minimum requirements or academic levels, as well as ensuring consistency, standardising (*"and to standardise any course content and teaching tools" [PTE]*), and ensuring the academic standard of these areas (*"to ensure that the curriculum that we teach meets the minimum requirements pertaining to the National standards of NCEA" [School]*).

Moderation-in-general was perceived to address the quality of student learning and achievement by nine respondents (two ITPs, seven PTEs). Moderation was seen to have a role in maintaining, monitoring, and ensuring consistency in student learning, progress, and achievement: *"[moderation functions to] ensure consistency of graduate outcomes" (ITP)*, and *"quality management of training outcome achievement" (PTE)*.

4.4.5.2 Internal moderation has a role in maintaining educational quality

The maintenance of educational quality was perceived as a role of internal moderation by 20 respondents (Table 4.26; 10% of qualitative sample). When commenting about internal moderation having this role, respondents also differentiated between programme, curriculum, and teaching aspects, and student learning and achievement aspects.

The role of internal moderation in maintaining the quality of the programme, curriculum, and teaching was raised by 17 respondents (two ITPs, 11 PTEs, four schools). As with moderation-in-general, this role included programme and course construction, teaching and teaching practices (*[internal moderation] is also about the learning and teaching process"* [ITP]; *"Peer interaction on education/teaching practices"* [PTE]), the review of teaching and curriculum, and resources and materials (*"Internal moderation should also provide the opportunity to review the learning and teaching materials"* [ITP]; *"[It ensures] that the delivery reflects the stakeholders' needs; all delivery materials and assessments are in-line with the learning outcomes"* [PTE]). Two respondents observed that, although internal moderation may provide information about programme, curriculum, and teaching, these aspects were not its primary focus.

The role of internal moderation in maintaining the quality of student learning and achievement was mentioned by 10 academic leaders (two ITPs, six PTEs, two schools). The PTE and school respondents endorsed internal moderation having this function. For example, *"The learner is at the heart of the learning and assessment process; public opinion does count, but probably not as important as the provision of student support and the quality outcomes the learner achieves" (PTE).* However, the ITP respondents were more reticent: one asserted that the role of internal moderation in programme, curriculum, and teaching was equally important to its role in student learning and achievement. The other opined that, depending on what the assessment involved covers, internal moderation may not be particularly effective in ascertaining students' readiness for employment (as had been suggested by item 9.3):

> *Moderation carried out will be dependent on programme structure and design, and unless there is a 'capstone' assessment covering all components of the graduate profile, it is difficult for moderation to actually determine whether students are prepared for employment or further study (ITP)*

4.4.5.3 NZQA moderation may not have a role in maintaining educational quality

NZQA moderation having a role in maintaining educational quality was mentioned by six respondents (Table 4.26; 3% of qualitative sample). Of these, only two PTE respondents endorsed this function: *"[NZQA moderation ensures that] quality teaching and assessment processes are evident" (PTE).*

Both ITP respondents qualified their endorsement of NZQA moderation having such a role. For example,

> *I would be concerned if external moderation reports were used as a significant factor in making judgements about the effectiveness of teaching and about learner achievement; however, they are one factor to take into consideration when making judgements about these two [key evaluative questions]. (ITP)*

The other respondents disagreed that NZQA moderation functions to maintain educational quality. For example,

> *[By implementing NZQA moderation feedback,] processes used within educational organisations to ensure a strong positive [educational] outcome can be so compromised that the result bears no relationship to the [educational] outcome for most of the students having been assessed (sic). [PTE]*

4.4.5.4 Discussion

The role of maintaining educational quality had immediacy for 15% of the qualitative sample: it was a function that sprang to mind for them when they initially thought of moderation-in-general. A further 18 respondents appear to have commented on (internal) moderation having this role after encountering closed-response items about internal moderation that addressed aspects of educational quality (e.g., curriculum: items 8.4 and 18.1, teaching: items 9.1 and 18.2, and student learning: items 9.2 and 9.3). That these 18 respondents had not mentioned educational quality beforehand suggests that the closed-response items may have prompted them to consider it as a function, and that they were sufficiently interested in it to comment. Lending weight to this inference is the fact that six of these comments appeared to directly respond to an item immediately preceding the open-field question. For example, *"Again, these statements [items 18.1 – 18.4] apply more to the validation or pre-assessment phase of moderation. At the post-assessment phase of moderation, they*

become issues that require urgent attention" (ITP). It is noted that most of those who did comment about internal moderation having a role in maintaining educational quality endorsed or elaborated on it: The tenor of most comments was positive.

The positive sentiment regarding moderation-in-general and internal moderation functioning to maintain educational quality was reflected in the quantitative data, where educational quality also featured: The factors, *Improving educational quality* and *Checking educational quality,* were extracted in the exploratory factor analyses as functions of internal moderation (see 4.1.1). The importance of the functions addressed by all associated items were relatively strongly endorsed: Three quarters or more of the respondents rated most as being of at least *Medium importance*, and 65% rated the other item as such. Quantitative analyses revealed that *Improving educational quality* and *Checking educational quality* were very strongly and positively correlated, indicating that respondents from PTEs and schools tended to rate the items associated with each the same way (see 4.1.4.1).

Quantitative analysis had revealed that the internal moderation survey items hung together in two factors (*Improving…*, and *Checking…*), but the strength of correlation between these two factors, and the qualitative data, suggested that these factors were closely related. Eight respondents (one ITP, three PTEs, four schools) referred to internal moderation having a role in improving educational quality (e.g., *"[Internal moderation is used] formatively to improve teaching and learning" [School]*), five (two ITPs, two PTEs, one school) referred to it checking educational quality, (e.g., *"[Internal moderation helps with] identifying gaps in teaching and learning" [School]*), and 13 (one ITP, 11 PTEs, one school) did not specify whether internal moderation *checked* or *improved* educational quality (e.g., *"There are key factors identified by internal moderation—teaching by the tutor and learning by the learners" [PTE]*).

The aspects of educational quality that respondents specified as being addressed by moderation (e.g., programme, curriculum, teaching, student learning) are reflective of findings in the literature that moderation can result in learning that is broader than being just about assessment. These nominated aspects align with the teaching and student learning parts of the teaching, student learning, and assessment cycle, reflecting the improvement impacts of moderation as found by some studies (see 2.4.4.2.2).

When the composition of the cohorts that referred to the maintenance of educational quality are considered in terms of organisation type (Table 4.26) in comparison to the composition of the qualitative survey sample (see Table 3.1), differences are noted for moderation-in-

general and for internal moderation.[116] Proportionally more PTE respondents referred to these roles, and proportionally fewer school respondents did. This suggests that when considering the main functions of moderation-in-general, the maintenance of educational quality was more likely to be 'top-of-mind' for those from PTEs than from schools, and that more PTE respondents felt the need to emphasise or expand on this function in their comments about internal moderation than school respondents did. These differences were reflected in the quantitative data about internal moderation. PTE respondents were found to have rated the items associated with *Improving educational quality* as having higher importance that school respondents did, and the items associated with *Checking educational quality* as having higher importance than either ITP or school respondents did (see 4.1.3). While the reasons for these differences are unclear, they may suggest that PTE respondents consider moderation to have a broader range of uses and functions than school respondents do. As one PTE respondent observed:

> *"To check that the teaching that students have received has adequately prepared them for assessment" [as per item 9.1] is an important aspect, but not as related to moderation of a single assessment and not truly part of moderation. But there must be a system within the organisation where the overall result of moderation, and a general review of assessments, is undertaken as a wide staff review. So, this is about the use of moderation results in reviewing teaching, self-assessment, and curriculum. (PTE)*

This broader range of uses of moderation reflects the calls in the literature for moderation to be a practice that focuses not only at 'one-point-in-time', but that addresses different stages in the teaching, student learning, and assessment cycle (see 2.4.0.1). The greater emphasis placed on the area of educational quality by PTE respondents as compared to those from schools may indicate that schools view moderation more narrowly than PTEs do, or that schools have other processes in place through which they maintain educational quality, which PTEs don't have. It may also be reflective of a business reality for PTEs that they use the tools they have available to ensure that their 'customers' are satisfied, as suggested by this response: *"By carefully analysing student results through the [internal] moderation, you can ensure that the student (customer) is gaining the knowledge that they have paid for and are expecting"* (PTE).

---

[116] No observations have been made about the composition of the cohort referring to NZQA moderation having this role due to the small cohort size.

Educational quality barely featured in the qualitative data about NZQA moderation, suggesting that few respondents considered it to be a function of, or to be emphasised by, this type of moderation. It is noted that educational quality was only a light focus in the NZQA Moderation section closed-response items, with one (item 21.5) concerning assessed content, and two (items 22.4 and 29.2) explicitly referring to teaching. Thus, it seems less likely that the survey contents directly prompted a response regarding this (as compared to the internal moderation section). However, respondents had been exposed to the internal moderation items prior to reaching this section of the survey, and thus, they had an opportunity to comment to this effect if they felt strongly that one of the functions of NZQA moderation was to maintain educational quality, or that NZQA moderation placed any emphasis on doing so. Educational quality did not appear in the quantitative data as a function of NZQA moderation (in that no *Educational quality* factor was extracted from the data, even though there were three items—as identified above—that would have made this possible). From the present qualitative and quantitative results, it appears that few academic leaders perceive NZQA moderation to emphasise or have a role in maintaining educational quality. However, to explore this more thoroughly, and to ascertain whether there were any observable differences in the perceptions of those from different organisation types, functions relating to educational quality would need to have featured more strongly in the closed-response items.

4.4.6 Dissatisfaction with NZQA

Dissatisfaction with NZQA was expressed by 23 respondents in their responses in the internal moderation or NZQA moderation survey sections, despite the survey items not soliciting such a sentiment (Table 4.27; 11% of qualitative sample). Respondents' dissatisfaction centred on three areas: NZQA moderation, NZQA assessment standards, and support and resourcing available. Dissatisfaction with NZQA did not explicitly feature in respondents' top-of-mind answers regarding the main functions of moderation, although two school respondents were cynical about the quality assurance function of moderation at the education system-level (see 4.4.3.1).

Table 4.27   Dissatisfaction with NZQA

	Overall		With NZQA moderation		With NZQA standards		With support & resourcing	
	n	%	*n*	%	*n*	%	*n*	%
Total	*23*		*19*		*7*		*7*	
ITPs	*2*	9	*2*	11	*0*	0	*0*	0
PTEs	*16*	70	*13*	68	*6*	86	*4*	57
Schools	*5*	22	*4*	21	*1*	14	*3*	43

Note: The overall totals may not equal the sum of the areas, as some respondents included more than area in their response.

4.4.6.1 Dissatisfaction with NZQA moderation

NZQA moderation was the focus of 19 respondents' dissatisfaction (refer Table 4.27). The ire of these academic leaders was focused on a range of aspects of NZQA moderation: process or requirements, feedback, purpose, NZQA moderators, perverse behaviours that occur in response, perceived inconsistencies, and the quality assurance of NZQA moderation itself.

NZQA's moderation process or requirements were the source of dissatisfaction for eight respondents (seven PTEs, one school). Complaints included the size of the assessed work sample submitted for moderation (*"… the sample is too small" [PTE]*), and the selection of that sample:

> *While it is easy to use pass [/] fail moderation results as an indicator of PTE practice, it is my belief that the system can be subverted through the selection of the best or most appropriate assessments [to be submitted] for moderation. I would be happier to have external moderation on scripts selected by the NZQA at 'random'. (PTE)*

Dissatisfaction with the number of standards moderated and with the approach taken to NZQA moderation also featured: *"Generally, I think the current system is weak and is not robust. There is no requirement for pre-assessment moderation [of assessment materials before use] …" (PTE)*.

The feedback provided in NZQA moderation reports was a source of dissatisfaction for five respondents (one ITP, two PTEs, two schools). Frustrations included that the feedback was lacking, of variable quality (*"The moderation reports are varied in terms of quality feedback …" [School]*), inconsistent with other advice received, or that it did not support improvement: *"[NZQA is] unlike some ITOs who give helpful and relevant [moderation] feedback to support*

best practice and provide details to support conversations as to why and how to improve"
(PTE).

Five respondents (two PTEs, three schools) were dissatisfied with the purpose of NZQA moderation. These respondents shared the perception that the moderation purpose was weighted towards accountability (and against an improvement of quality), and they railed against this weighting. For example, *"[The purpose of NZQA moderation is] to make cabinet happy rather than genuinely look to improve teaching and learning" (School)*; and

> *[There was a] lack of input [or] contributions from NZQA at the Moderation process stage; [it was] totally focused on [the moderation] outcome, rather than how to add more value into the process which can lead to a better outcome. (PTE)*

Five respondents (four PTEs, one school) expressed dissatisfaction with the NZQA moderators. The dissatisfaction of the PTE respondents centred on a perceived lack of industry experience or adult teaching experience on the part of the moderators. For example, *"[Any] quality assurance functions [of] moderation by NZQA are limited by the lack of 'real' experience moderators have [had of] operating their knowledge in a 'real' world with the realities that businesses and workplaces face" (PTE).*

That NZQA moderation can lead to a range of perverse behaviours within organisations was of concern to four PTE respondents. The perverse behaviours suggested ranged from manipulating the samples submitted for moderation (as per the above quote referring to subverting the system), to taking approaches that compromise learning (e.g., *"Most people I talk to think NZQA promotes an inflexible approach that inhibits learners learning the fundamentals of efficiency and common sense" [PTE]*).

Three PTE respondents perceived there to be inconsistencies in the NZQA moderation requirements or between NZQA moderators: *"The issue I see … is that when we send off our [moderation submission], [what the moderation result will be] hangs totally on the person moderating the marking at the time" (PTE).* The issue of quality assurance of NZQA moderation itself was of concern to a further two respondents (one ITP, one PTE): *"External Moderators require quality assurance too" (PTE).*

4.4.6.2 Dissatisfaction with NZQA assessment standards

Dissatisfaction with the NZQA assessment standards was the second main area of concern, as expressed by seven respondents (Table 4.27). Most (five PTEs) suggested that some standards were not fit-for-purpose. Three of them took issue with the requirements of unit standards in the Core Health (First Aid) field. The other two felt that the requirements of some standards lacked clarity (*"[It] concerns me that there [were] not enough clear guidelines in the unit standards at the very beginning to ensure consistency" [PTE]*). Revision of the standards was of concern for three respondents (two PTEs, one school). For example,

> *NZQA moderation would benefit from using the feedback to improve a unit standard if this no longer meets the industry standard, i.e. is not fit-for-purpose. … NZQA departments seem to be a bit separated, so the standards can lag behind best practice, [and] even if the moderators realise it, they still moderate specifically to the unit standard [evidence requirements]. (PTE)*

4.4.6.3 Dissatisfaction with support and resourcing available

Support and resourcing in the education sector was a third focus of dissatisfaction, with seven respondents (Table 4.27) expressing this sentiment. Among these respondents, the amount of, or access to, support (e.g., assessment exemplars, training, advice) was a source of concern: *"NZQA do not (sic) provide pre[-assessment] moderation [of assessment materials] or any forum to share and/or debate best practice, or assessment for that matter" (PTE)*. Respondents also felt that more resourcing (in terms of time and money) was required:

> *To ensure a high level of public confidence the government would need to invest: more time to teachers for marking, exemplars that have been approved and can be utilised with very little effort by teachers, … If [a focus on educational quality] is the intention of internal moderation, it should be better funded, and training and support made available to schools free of charge, e.g., best practice workshops (originally free, now with a cost). (School)*

4.4.6.4 Discussion

Dissatisfaction with NZQA was not solicited by the survey items. That 11% of respondents expressed this sentiment suggests that they felt strongly enough to take the opportunity to

record their dissatisfaction, although it was not elicited. Some respondents may have also felt encouraged to do so by the knowledge that at the time of data collection, the researcher was employed by NZQA, although it was explicitly stated that she was conducting the current study as a private citizen. Nonetheless, some respondents may have recorded their dissatisfaction with NZQA in the hope that the researcher would relay this to her employer. (It is noted that this did not occur.)

Each of the three main areas of dissatisfaction (NZQA moderation, NZQA standards, support and resourcing) featured as a concern for respondents from both the secondary and tertiary sectors. It is unlikely that the organisations of all respondents expressing dissatisfaction were moderated for the same standards. The issues do not appear to have been limited to one group of standards (e.g., English achievement standards or Core Health unit standards), one type of assessment standard (i.e., achievement standards or unit standards), or to the moderation conducted by one NZQA business unit or another. Instead, it suggests that the issues causing concern were relatively widespread (refer 1.2.1).

When the composition of the cohort who expressed dissatisfaction with NZQA is considered in terms of the organisation types of respondents (Table 4.27), it is evident that fewer schools and more PTEs expressed dissatisfaction as would proportionally reflect the composition of the full qualitative sample (Table 3.1). The reasons for this are unclear but could include contextual differences between the secondary and tertiary environments (e.g., the differences in the nature and specificity of requirements placed on the different organisation types regarding internal assessment by NZQA; see 2.4.3.2.1). Other differences between the secondary and tertiary sectors could also be implicated, such as the differences in NZQA's role as a quality assurance body for organisations in each (MNA visits for schools, EER for tertiary organisations; see 2.3.1.2, 2.3.1.3.1), and potential differences in communications channels between NZQA and the different types of organisation. Other reasons could also include differences in the standards that each organisation type tends to assess against (see 1.2.1). It seems likely that a contributing factor to the level of dissatisfaction expressed by PTE respondents were the Core Health unit standards, which, anecdotally, were contentious. At the time of data collection, NZQA was the standard-setting and quality assurance body responsible for the Core Health unit standards, and as such, managed their national moderation. PTEs were the main users of these standards; some PTEs only delivered First Aid training. Thus, the negative sentiment and dissatisfaction with NZQA expressed by some PTE respondents was likely exacerbated by their frustration regarding the Core Health unit standards, and NZQA's moderation of those standards.

4.4.7 Qualitative results: Summary

In conclusion, academic leaders were found to believe that moderation-in-general and internal moderation function in five very similar areas, and that NZQA moderation functions in four of those areas. Quantitative and qualitative findings largely aligned and complemented each other; each provided nuance to the findings of the main areas that moderation was seen to function in. Furthermore, quantitative analyses established that for most functions, there were no real differences in how academic leaders from the different organisation types saw moderation. For the few functions where observable differences were found, PTE respondents tended to rate moderation as having a higher level of importance or emphasis than those from the other organisations. In the next chapter, these results are used to answer the present study's research questions.

5. Discussion

This study originally set out to answer two research questions:

1. What do academic leaders in ITPs, PTEs, and schools, perceive the *in-practice* functions of internal moderation and NZQA moderation to be?

2. To what extent are there differences observable between the three organisation types regarding academic leaders' perceptions of the *in-practice* functions of internal moderation and NZQA moderation?

These two questions are about moderation *in-practice*: what actually occurs. However, in keeping with the nature of a mixed-methods pragmatic approach, inductive analysis of the qualitative survey data revealed that a *post hoc* research question was also raised and answered:[117]

- What do academic leaders in ITPs, PTEs, and schools, perceive the *in-principle* functions of moderation to be?
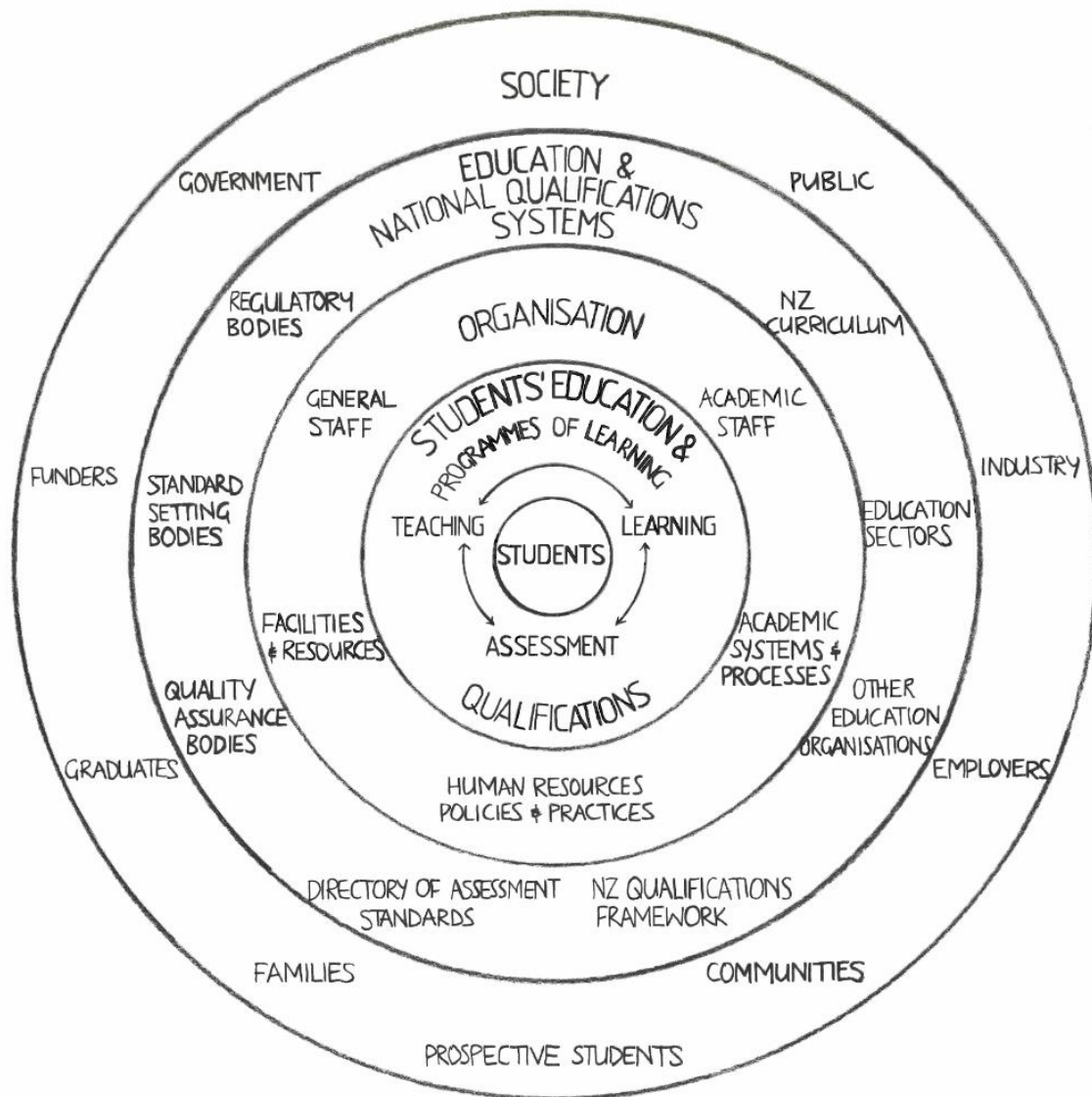
Firstly, a contextual framework for moderation is introduced which will assist with answering the research questions. The *post-hoc* question (about *in-principle* functions) is then answered, before the original *in-practice* questions are addressed. Links to the literature are made as the *post-hoc* question is answered but are not repeated in the subsequent answers.

5.1 Contextual framework for moderation

A contextual framework for moderation (visually represented in Figure 5.1; clean template included as Appendix 17) has been developed for use as an analytical tool to assist with making sense of the results of the present study. In this framework, the multiple contexts within which the perceived roles and functions of moderation occur are depicted as a set of concentric circles, representing layers of embedded contexts (refer 1.2.5). The context immediate to internal assessment (as the concern of moderation) is embedded within a series of increasingly broader contexts that are further removed from internal assessment events. Interaction occurs between contexts (shown as adjacent layers), and in the form of driving influences exerting force from super-ordinate to subordinate contexts (i.e., across the layers towards the centre of the framework).

---

[117] This occurred through the data and analyses pertaining to the results reported for moderation-in-general in 4.4. Refer 3.4.5.1.

Figure 5.1   Contextual framework for moderation

The outermost layer of the framework is the overarching context of NZ society, comprising the government, funders, industry and employers, and the public. The local and wider communities, parents and families, prospective students, and graduates also form part of this layer. The broad policy directions as issued from central government emanate from this layer, and drive or influence what occurs in each of the subordinate layers (contexts) of the framework. The societal layer wraps around the other components of the structure, reflecting what is arguably one of the fundamental purposes of education and training, that of preparing and equipping students to participate in wider society (European Association for Quality Assurance in Higher Education et al., 2015; Morshead, 1995).

The context of the education and national qualifications systems sits adjacent and internal to the societal layer in the framework. This layer comprises the different sectors of the NZ education system, the NZ Curriculum and other curricula relevant to the sector, and the NZQF and DAS (see 1.2.1). The education and national qualifications systems layer also contains the quality assurance, regulatory, and standard-setting bodies (including the Ministry of Education, NZQA, and the Education Review Office). The policy directives from these bodies (e.g., the Ministry or NZQA), the external requirements that are imposed on organisations, and the external quality assurance framework within which organisations operate, are parts of this layer of the framework, and influence (or exert a level of control over) the subordinate layers. Other education organisations also form a part of this contextual layer.

The organisation context layer is embedded within the education and national qualifications systems context. It encompasses all aspects of an education organisation, including personnel (general or non-academic staff, and academic staff including teachers, assessors, and academic leaders), academic systems and processes (e.g., development, quality assurance including moderation, and review of programmes of learning, and assessment), and human resources policies and practices (e.g., professional development). Facilities, resources, and services (e.g., those providing student support) are also included. In New Zealand, there is likely to be a degree of variation in the characteristics of this layer as it applies to different individual organisations, due to most being self-managing (see 1.2.4). This layer provides the context for the adjacent internal layer—that of students' education and qualifications.

Students' education and qualifications is the context forming the immediate layer surrounding the framework's core (students). This context comprises the education, programmes of learning, and qualifications that students engage in and receive; the teaching, learning, and assessment cycle is integral to this layer of the framework. The taught curriculum, and teaching practice, activities, and resources that students experience are encapsulated within the layer. Students' learning and (formative and summative) assessment experiences, and their progress and achievement, also form part of this contextual layer.

Students sit at the core of the framework, in recognition of their centrality to education, the education system, and to education organisations that deliver education and assess for qualifications. At a certain point in time, most students exit formal education and move from occupying the central core to occupying the structure's outer layer, as members of society.

This depiction emerged from the process of the researcher trying to make sense of the results of this study and was inspired by qualitative data from both stages of the study (interviews and survey). Both data sets contained evidence of perceptions of differing breadths of moderation focus (from narrow and directly focused on assessment, to broad and considering peripheral matters, e.g., professional development). Further, both contained implicit references to different structural levels (e.g., education-system level: *"ensure consistency across New Zealand"* and *"parity across the system"*; individual organisation level: *"all staff in a department"*, *"compliance across the organisation",* and *"when issues are raised, the trainer failure is addressed"*; and specific, individual assessment instances: *"the consistency of judgements between different students' evidence"*). Thus, although academic leaders did not explicitly refer to the layers of embedded context within which they perceive moderation to function, their responses alluded to these contexts.

5.2 What academic leaders perceive the functions of moderation to be, in principle

The present study found that academic leaders perceived that, in principle, moderation can function in five areas: *assessment quality, educational quality, professional and organisational learning and development, organisational and wider quality assurance,* and *public and stakeholder confidence.* This finding differs from what was found in the literature review (2.4.3–2.4.6), as discussed below. The five areas identified in the present study will be amplified in sections 5.2.1–5.2.5. They can be interpreted using the contextual framework for moderation (Figure 5.1). Figure 5.2 shows these *in-principle* functions overlaid on the framework; the areas in which moderation was seen to simultaneously function sit in different layers. The areas differ in breadth of focus. The narrowest function relates directly and solely to the quality of internal assessment. The functions that consider, inform, or respond to aspects that are adjacent, peripheral, or indirectly related to internal assessment (e.g., as part of the context within which assessment occurs) are referred to as broad because they encompass more than internal assessment alone.

Much of the current literature holds that there are two fundamental groupings of purposes of moderation: those associated with accountability and quality control, and those associated with improvement and learning (e.g., Ehren et al., 2013; Klenowski & Wyatt-Smith, 2014; see 2.4.3, 2.4.4). These two groups echo the distinction between summative and formative uses of assessment (see 2.1.1; e.g., Biggs & Tang, 2007; Broadfoot, 2007; Crooks, 2011). While

some of the *in-principle* functions of moderation, as perceived by academic leaders and found in the present study, appear on the surface to align more closely with one grouping or the other, others do not. Further, those that appear to align more closely with one also appear to support the other. This lack of clear alignment perhaps supports the critiques of the notion of an accountability-versus-improvement dichotomy (e.g., Ehren & Hatch, 2013; Harvey & Newton, 2007; see 2.4.5). Adie et al. (2013) identified more nuanced understandings of moderation than the dominant discourses of accountability and improvement (see 2.4.6). The different understandings of moderation as per Adie and colleagues align in part, but not entirely, with the *in-principle* functions of moderation as found by the present study. These alignments and cross-overs are identified as the different *in-principle* functions are discussed below.
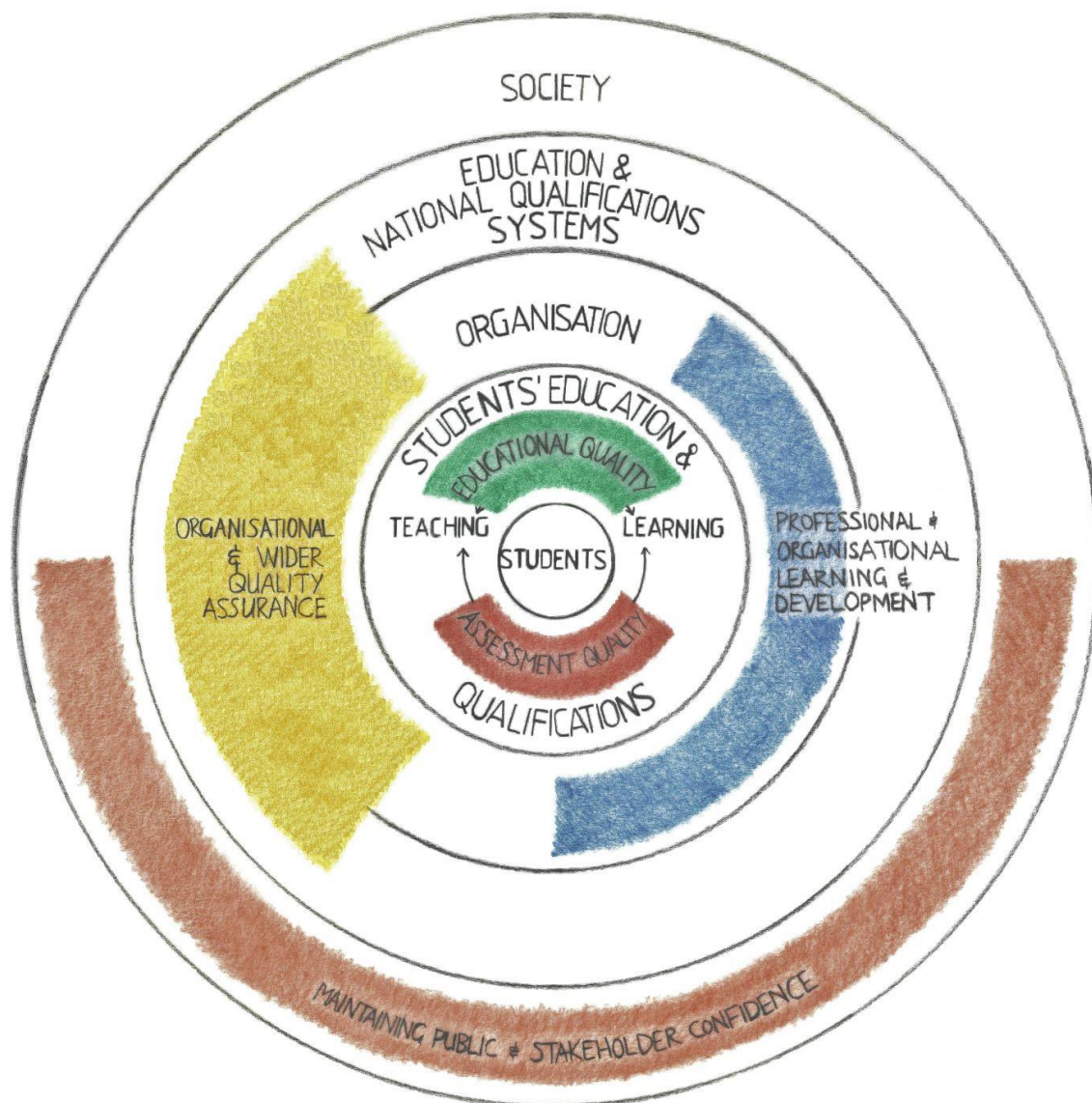
Figure 5.2   Perceived *in-principle* functions of moderation

5.2.1 Assessment quality

The narrowest conception and use of moderation as perceived by academic leaders pertains to the quality of internal assessment (Figure 5.2). This use featured most strongly in the data (see 4.4.1.1), which is unsurprising, as moderation is usually defined as a process through which internal assessment is quality assured (e.g., Harlen, 2007).

Moderation was perceived to have different actions on assessment quality: that of assuring, evaluating (i.e., quality control), and of providing feedback and improving (see 4.4.1.1.2 and Table 4.21). The evaluation action aligns partially with Adie and colleagues' (2013) notion of moderation as accountability (refer 2.4.6), in that both consider the grades awarded. The evaluation and improvement actions reflect the dominant discourses of accountability and improvement (see 2.4.3 and 2.4.4). However, the strong presence of the assurance action (which is broader than either and encompasses both evaluation and improvement) suggests that when academic leaders think about the *in-principle* function of moderation that pertains to assessment quality, most do not distinguish between control and improvement; instead they appear to think of the actions of moderation regarding assessment quality in a more holistic, encompassing, or overarching way. The prevalence of the assurance action in the present findings coheres with Adie and colleagues' (2013) discourse of moderation as equity, where moderation functions to ensure that assessments are sound and trustworthy (see 2.4.6).

Moderation was also perceived to address different components of the assessment process (4.4.1.1.1, Table 4.20)—different links in Crooks and colleagues' (1996) chain (see 2.1.3 and 2.4.0.1)—including the assessment instrument, practices and processes, and assessor judgements, as Bloxham et al. (2016) had also found. In addition, this study found academic leaders perceived that, in principle, moderation considers a comprehensive range of properties of internal assessment (e.g., that what is meant to be assessed is assessed at the correct academic level, the assessment is fair, and assessor judgements accurate; refer 4.4.1.1.3 and Table 4.22). When taken together, these properties represent the technical quality of assessment in Newton and Shaw's (2014) evaluative framework (see 2.1.3). These properties also are found in the conception of moderation as equity, and to a lesser extent (i.e., primarily the property of accuracy of assessor judgements) in the conception of moderation as justification (Adie et al., 2013; refer 2.4.6).

Moderation, when focused on assessment quality, is a lens through which it is possible to consider internal assessment in isolation. When used and conceptualised in this way,

moderation can have a narrow focus that ignores the context within which assessment occurs. However, if the programme of learning, and the teaching, student learning, and assessment cycle within which assessment occurs are ignored, assessment activities and practices that are fragmentary and that do not support or enhance learning might be allowed to continue unchecked (Altrichter & Kemethofer, 2015; Biggs & Tang, 2007; Black et al., 2011; Broadfoot, 2007; Crooks, 2011; Ehren & Hatch, 2013; Ehren et al., 2015; Gustafsson et al., 2015; Harlen, 2007; Miller et al., 2009).

5.2.2 Educational quality

The maintenance of educational quality is a broad *in-principle* use of moderation as perceived by academic leaders that focuses on aspects of the students' education and qualifications layer of the contextual framework for moderation (Figure 5.2; refer 4.4.5.1 and Table 4.26). This conception of moderation recognises assessment as being situated and considered in the context that surrounds assessment. It affords a view of a programme of learning or course within which assessment occurs, and the associated teaching and pedagogy. In addition, it affords a view of students' learning progress (reflecting one of the aims of Black et al., 2010, 2011, in their longitudinal intervention; see 2.2.3.1, 2.4.4.2.2), and whether students are adequately prepared for their future endeavours—an aspect that references one of the fundamental purposes of education (e.g., European Association for Quality Assurance in Higher Education et al., 2015; see 2.1). Thus, moderation provides a lens through which to consider the programme of learning and other parts of the teaching, student learning, and assessment cycle (e.g., Beutel et al., 2017; Harlen, 2007; Miller et al., 2009; see 2.4.4.2.2).

In considering aspects that pertain to educational quality (such as whether assessment supports student learning, i.e., the impact of assessment), this *in-principle* function of moderation provides a mechanism through which to evaluate aspects of the social value of assessment and assessment systems, as per Newton and Shaw's (2014) evaluative framework (see 2.1.3). When the data relating to this function of moderation were considered in terms of the purposes of accountability or improvement, it became apparent that the function relates, in part, to both accountability and improvement, but aligns cleanly with neither. This suggests that, as for the function of assessment quality, when thinking about the role of moderation in maintaining educational quality, academic leaders tend to consider it in a holistic and encompassing way (as opposed to drawing a fine distinction between the accountability and improvement aspects).

5.2.3 Professional and organisational learning and development

Another *in-principle* function of moderation as perceived by academic leaders pertains to professional and organisational learning and development (refer 4.4.4.2 and Table 4.25). When interpreted through the contextual framework, professional and organisational learning and development is a broad conceptualisation of moderation that focuses at the organisation layer, in that it pertains to academic, personnel, and organisational processes, practices, and procedures (Figure 5.2). This conception recognises that moderation can provide an overview of professional and organisational learning and development needs and can offer professional and organisational learning and development opportunities, much as concluded by Adie (2012), and Misko (2015b)—see 2.4.4.2; 2.4.5.

Academic leaders in the present study considered that the professional learning and development opportunities that moderation could provide pertain to any aspects of the teaching, student learning, and assessment cycle (refer 4.4.4.2). Academic leaders identified that these opportunities could be about teaching and student learning (e.g., teaching content), assessment (e.g., design, methodology, practice, judgements), or processes such as moderation. The range of aspects addressed by these opportunities aligns with the findings of multiple studies (e.g., Black et al., 2010; 2.4.4.2.2). Opportunities for developing shared understandings between people about assessment and the requirements of standards can also be provided, in line with previous findings (e.g., Watty et al., 2014; 2.4.4.2.4). Furthermore, moderation can support communities of practice, which also provides learning and development opportunities, reflecting others' findings (e.g., Crisp, 2017; 2.4.4.2.10). The conception of moderation as community building (Adie et al., 2013; 2.4.6) aligns closely with these aspects of this *in-principle* function of moderation.

Academic leaders in the present study tended to believe that, in principle, moderation can also provide opportunities for organisational learning and development relating to academic and quality assurance practices and processes including assessment and moderation. These opportunities echo others' findings (e.g., Grant, 2012; 2.4.4).

The *in-principle* moderation function of professional and organisational learning and development pertains to improvement and learning, by definition. However, it also may result in improvements in quality, and as such, greater consistency, reliability, and compliance with requirements, as found by O'Connell et al. (2016; 2.4.3.4), and asserted by Harvey and Newton (2007; 2.4.5). Therefore, through these impacts, this conception of moderation may also be seen to support accountability and quality-control purposes.

5.2.4 Organisational and wider quality assurance

Moderation is also seen by academic leaders as serving quality assurance in areas other than assessment (refer 4.4.3.1, Table 4.24). This broad conceptualisation, when considered in terms of the contextual framework, addresses the organisation layer, and education and national qualifications systems layer (Figure 5.2).

In this conception, moderation functions as a lens through which other organisational processes, practices, and factors can be viewed. As such, it specifically serves organisational quality assurance. The academic leaders in the present study tended to believe that, in principle, moderation could provide information to inform various organisation-level processes, including those that only indirectly impact on assessment (e.g., organisational self-assessment or evaluative review processes). Moderation was also seen to meet other quality assurance requirements, such as enabling ongoing compliance with accreditation or consent to assess requirements, adherence to an organisation's own quality management system requirements, or ensuing that external requirements are complied with. These uses reflect the intentions of many quality assurance frameworks (e.g., Croxford et al., 2009; Ehren et al., 2013; Misko, 2015a; see 2.3.0.2).

Moderation was also seen by academic leaders to function at an education system level, to several ends. It was recognised as a medium through which quality assurance bodies or governance regimes can ensure that organisations comply with external requirements, a use that reflects quality assurance functions as identified by others (e.g., Altrichter & Kemethofer, 2015; Misko, 2015a, 2015b). Further, academic leaders saw *in-principle* functions of moderation as being to quality-assure qualifications at the education system level and to ensure comparability across the sector. These perceptions echo others' sentiments regarding the fundamental purpose of education system quality assurance regimes (e.g., Ehren et al., 2013; 2.3).

Most aspects of the *in-principle* function of organisational and systems-level quality assurance, as expressed by academic leaders, appear to have quality control and accountability foci. As such, they work to enable stakeholders who are internal and external to the organisation, as well as the public, to have confidence in the education organisation, the qualifications it awards, and the education system as a whole (see 2.4.3). Thus, conceptually, the *in-principle* function of organisational and wider quality assurance enables the maintenance of public and stakeholder confidence (European Association for Quality Assurance in Higher Education et al., 2015; 2.4.3). However, the organisational and wider

quality assurance function could also support improvement purposes, for example, through the actions associated with accountability-focused quality assurance processes that can result in improvements occurring (as found by Ehren & Hatch, 2013; see 2.4.5). The finding that *in practice*, the functions of organisational quality assurance and maintaining public confidence are only moderately correlated (refer 4.1.4.1, Table 4.10; 4.2.4.1, Table 4.16) lends weight to the caution that alignment with one purpose does not preclude alignment with or reinforcement of the other.

5.2.5 Maintaining public and stakeholder confidence

Maintaining public and stakeholder confidence, a societal-layer function, was the fifth area identified by academic leaders as being an *in-principle* function of moderation (refer 4.4.2.3 and Table 4.23, Figure 5.2). Graduates, parents, and employers were among the stakeholders identified by academic leaders as those whose confidence was to be maintained (Figure 5.1). These reflected the stakeholders that Gilbert (2012) had identified in relation to education organisations (see 2.4.3.1). When considered in terms of Newton and Shaw's (2014) evaluative framework, the function of maintaining public and stakeholder confidence is closely tied to the social evaluation of assessment: the acceptability of implementation (see 2.1.3). In terms of Adie and colleagues' (2013) conceptualisations, moderation as justification holds some similarities, in that moderation enables justification of assessor judgements to various stakeholders if needed, which in turn should maintain those stakeholders' confidence (see 2.4.6).

Fundamentally, this *in-principle* function is about maintaining confidence in the education and qualifications provided by an organisation and ensuring that these are credible (e.g., Controller and Auditor-General, 2012; see 2.3, 2.3.0.2; 2.4.3). The education and qualifications provided by an organisation encompasses programmes of learning, the entirety of the teaching, student learning, and assessment cycle, student achievement, and graduate knowledge and competence. However, as Harlen (2007) observed, the public and stakeholders tend to have lower confidence in internal assessment than they do in external assessment. As such, it appears that maintaining confidence in the *internal assessment conducted* in an organisation can sometimes become a proxy for maintaining confidence in the *quality of education and qualifications delivered*. When this is so, the function of moderation maintaining confidence narrows in focus. The maintenance of public and stakeholder confidence is related to accountability and quality control functions and foci, in that accountability and quality control serve the maintenance of confidence (e.g., European

Association for Quality Assurance in Higher Education et al., 2015; 2.4.3.1). However, serving accountability-purposes does not preclude also serving improvement-purposes. Although the volume of data regarding this *in-principle* function was small, there was no indication in the data that academic leaders perceived the maintenance of public and stakeholder confidence to only be for accountability purposes. Instead, the language used was primarily of 'assurance', suggesting again a more holistic and encompassing conceptualisation than breaking the function down to checking (control) versus improvement.

# 5.3 What academic leaders perceive the functions of moderation to be, in practice

5.3.1 Internal moderation, in practice

Academic leaders were found to believe that, in practice, internal moderation functions in five main areas: *assessment quality*, *maintaining public and stakeholder confidence*, *organisational quality assurance*, *professional and organisational learning and development*, and *educational quality* (Figure 5.3). For the most part, these areas are the same as the *in-principle* functions of moderation, except for *organisational quality assurance*, which was narrower in scope than the *in-principle* function of *quality assurance*. This study found that internal moderation is perceived to work simultaneously in these areas, which occupy different layers of the contextual framework.

The five areas in which internal moderation was seen to function in practice differ in breadth of focus. Academic leaders appear generally to perceive that, in practice, internal moderation has narrow functions that focus directly on the quality of internal assessment, and broad functions that focus on teaching and student learning, along with aspects that occur across the other layers of the conceptual framework (similar to their perceptions of the *in-principle* functions of moderation; 5.2). The correlations between the internal moderation scales lend weight to this apparent distinction between narrow and broad functions: the scales pertaining to educational quality, organisational quality assurance, professional learning opportunities, and organisational development are all strongly and positively correlated with each other (see 4.1.4.1, Table 4.10). *Maintaining public confidence* is also moderately positively correlated to these scales (at an aggregated organisation level), and for PTEs, strongly positively correlated to most. In contrast, the one internal moderation

scale pertaining to assessment quality is only moderately positively correlated (at best) with the scales addressing the broader functions of internal moderation.

Academic leaders' perceptions of the *in-practice* functions of internal moderation pertaining to assessment quality did seem to distinguish between accountability and improvement purposes (this is expanded on in 5.3.1.1). However, respondents did not appear to make this distinction in the other four areas in which they saw internal moderation to function.
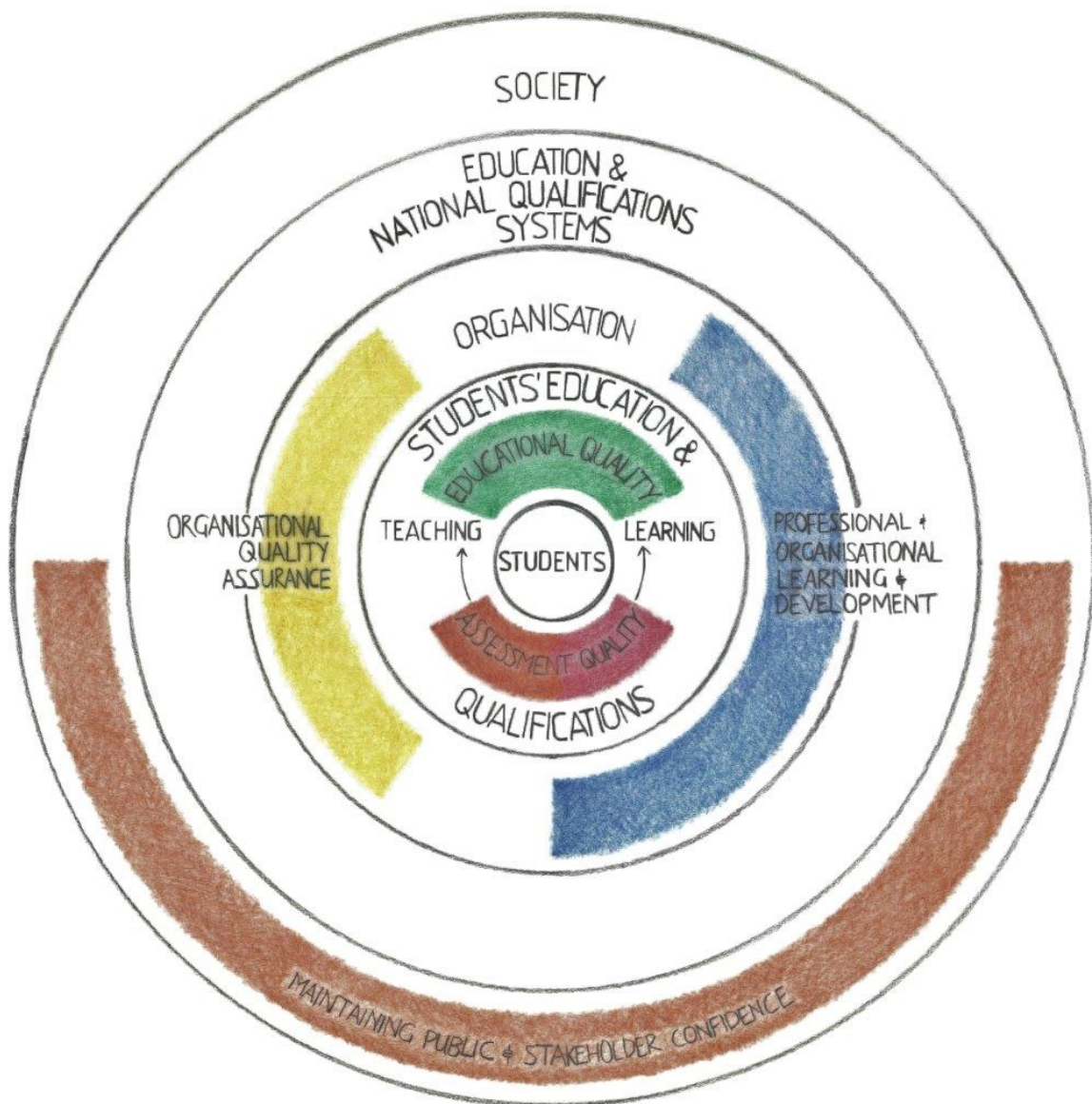
Figure 5.3   Perceived *in-practice* functions of internal moderation

5.3.1.1 Assessment quality

The narrowest area in which internal moderation was perceived to function was assessment quality, endorsed by all respondents (Figure 5.3). How academic leaders saw internal moderation in regard to assessment quality *in-practice* was nuanced: *Checking assessor judgement quality*, *Checking assessment material quality*, and *Improving assessment material and assessor judgement quality* factors were extracted in the quantitative analysis (refer 4.1.1). The quantitative data suggest that academic leaders saw 'checking' quality as distinct from 'improving' quality, as evinced by items pertaining to each being captured by separate factors, and by the level of importance ascribed to the items associated with the factors—although the qualitative data did not reinforce such a clear distinction. Respondents almost unanimously rated the items associated with *Checking assessor judgement quality* and *Checking assessment material quality* factors as having high importance; the unanimity and homogeneity of response were so great that the items associated with these factors could not be calibrated onto measurement scales. In contrast, the level of importance ascribed to the items associated with the factor *Improving assessment material and assessor judgement quality* varied (although three quarters of respondents strongly endorsed the items; see 4.1.1). Assessment quality was the only area of the *in-practice* functions of internal moderation that clearly aligned with one or other of the accountability-quality control versus improvement purposes as features in current literature (refer 2.4.3, 2.4.4, 5.2).

The quantitative data also suggested that academic leaders tend to see the *in-practice* functions of internal moderation pertaining to assessment quality as relatively 'stand-alone'. Although no correlations could be calculated for the 'checking' of assessment quality aspects, *Improving assessment material and assessor judgement quality* was found to only be weakly or moderately (positively) correlated with other internal moderation functions (see 4.1.4.1, Table 4.10). Internal moderation was rated as having higher importance than the perceived level of emphasis of NZQA moderation on all items addressing assessment quality that were used to probe respondents' views of both moderation systems (see 4.3.2; Table 4.18).

Assessment quality featured strongly in the qualitative data pertaining to the *in-practice* functions of internal moderation (refer 4.4.1.2). The component of the assessment process that internal moderation addresses, the action that internal moderation takes on that assessment component, and the properties of assessment with which internal moderation is concerned were the aspects to which academic leaders referred in their responses. Internal

moderation was variously seen to assure, evaluate, and provide feedback about and improve assessment-in-general (including assessment practices and processes), assessor judgements, and assessment materials. Respondents indicated that internal moderation is concerned with the extent to which what was intended to be assessed, was in fact assessed at the required level, and whether assessments were fair, of robust quality, and consistent (in terms of assessor judgements, comparability of grades awarded, and between organisations). It seems likely that respondents also considered internal moderation to be concerned with the accuracy of assessor judgements.[118]

5.3.1.2 Maintaining public and stakeholder confidence

Maintaining public and stakeholder confidence was another *in-practice* function of internal moderation identified in the data. This is a broad function that acts at the societal layer of the contextual framework (Figure 5.3). While accountability and quality control enable the maintenance of public and stakeholder confidence (2.3.0.2, 2.4.3.1), the correlations in the quantitative data did not suggest that maintaining public confidence was clearly aligned with an accountability or control agenda. Instead, this function was at least moderately positively correlated (for aggregated organisations) with all the other functions of internal moderation, and for PTEs, and strongly so with some (see 4.1.4.1, Table 4.10).

The maintenance of public and stakeholder confidence was identified in both data sets. The items associated with the factor *Maintaining public confidence* (which addressed the maintenance of public confidence in national qualifications, and in the internal assessment, education programmes, and qualifications of organisations) were strongly endorsed by respondents (see 4.1.1).

Maintaining public and stakeholder confidence as an *in-practice* function of internal moderation also featured strongly in the qualitative data (see 4.4.2.1). Respondents referred to the public and stakeholders as comprising a wide range of direct and indirect stakeholders external to (i.e., not employed by) the organisation, most of whom (apart from current students), sit in the societal layer of the contextual framework (Figure 5.3). Most respondents who referred to this function broadly endorsed internal moderation as maintaining public and stakeholder confidence. Various perspectives were put forward, including that, in practice,

---

[118] The property of accuracy of assessor judgements was included among those that moderation-in-general was seen to be concerned with (refer 4.4.1.1), and although not explicitly identified as being of concern to internal moderation, the quantitative data indicated that most respondents saw it as having high importance as a function of internal moderation (see Figure 4.1).

the role of internal moderation in maintaining public and stakeholder confidence was indirect (through being a part of an organisation's quality management system), and that internal moderation lacks public visibility. The lack of visibility appeared to justify some respondents' disregard for this function (as they therefore presumably equated it to lacking effectiveness in maintaining public confidence). However, other respondents appeared to believe that public and stakeholder confidence was served by the knowledge of the existence of quality assurance processes such as moderation, rather than knowledge of how such processes work.

5.3.1.3 Organisational quality assurance

Organisational quality assurance was another area in which internal moderation was seen by academic leaders to function in practice. This quality assurance pertains to the organisational layer of the contextual framework (Figure 5.3), and as such, is more limited in scope than the *in-principle* quality assurance function of moderation.[119] Organisational quality assurance is a broad conception of moderation: Academic leaders believed that in practice, internal moderation has quality assurance functions relating to (or a part of) the organisational systems, processes, and requirements that encompass aspects beyond the quality of internal assessment alone. It appears that academic leaders tend to see the organisational quality assurance function as part of the broader set of functions of internal moderation, more so than aligning with accountability or improvement purposes. The strength of correlations between the organisational quality assurance scale and those addressing educational quality, professional learning opportunities, organisational development, and maintaining public confidence (4.1.4.1, Table 4.10) lend weight to this inference. Organisational quality assurance was identified as an *in-practice* function in both quantitative and qualitative data sets.

Four of the five quantitative items associated with the *Organisational quality assurance* factor (pertaining to supporting organisational self-evaluation, meeting internal and NZQA requirements, and providing confidence to organisational management) were strongly supported by respondents (see 4.1.1, Figure 4.5). In contrast, the item that suggested internal moderation provides evidence for performance appraisals was relatively weakly endorsed. Even so, internal moderation was rated by academic leaders as having higher

---

[119] Which also included aspects in the education system context layer (see 5.2.4, Figure 5.2).

importance for providing evidence for such appraisals than the level of emphasis that NZQA moderation was believed to have (see 4.3.2).

The organisational quality assurance function of internal moderation also featured in the qualitative data (see 4.4.3.2). Beliefs expressed by respondents included that internal moderation is a part of an organisation's quality management system, is used to inform other organisational processes, is an externally set requirement, and has a role in confirming quality.

5.3.1.4 Professional and organisational learning and development

Professional and organisational learning and development was also seen by academic leaders as an *in-practice* function of internal moderation (Figure 5.3). This is also a broad area that does not directly address internal assessment quality, but instead can impact on any aspect of the teaching, student learning, and assessment cycle (or other target areas). This conception of moderation sits at the organisation layer of the contextual framework and pertains to the learning and development of the personnel of an organisation, and of the organisation itself. Although, by definition, this *in-practice* function of internal moderation aligns with improvement and learning purposes (see 5.2.3), the correlations between the two quantitative factors (refer below; and see 4.1.4.1, Table 4.10) and the other internal moderation factors do not suggest that such an alignment is exclusive. Instead, the correlations suggest that this *in-practice* function is aligned with both accountability and improvement purposes, and that academic leaders usually see it as part of the broad-focused functions of internal moderation. Professional and organisational learning and development featured in both data sets.

Providing professional learning opportunities and *Assisting organisational development* were the two relevant factors extracted from the quantitative data (see 4.1.1). Respondents strongly endorsed the importance of most items associated with these two factors, and endorsed the balance as having medium or high importance (see Figures 4.6 and 4.7). *Providing professional learning opportunities* and *Assisting organisational development* were found to be strongly and positively correlated for all organisation types (see 4.1.4.1); these were the strongest correlations between internal moderation scales. This lends further weight to the aspects of internal moderation aligned with these factors being a part of the larger area of professional and organisational learning and development. Internal moderation was rated as having higher importance than the level of emphasis that NZQA moderation

was perceived to have for the functions of developing shared understandings among people and building confidence (see 4.3.2).

The role of internal moderation in professional and organisational learning and development was also emphasised in the qualitative data (see 4.4.4.1). Specifically, respondents saw internal moderation as offering opportunities for professional learning and development, contributing to communities of practice, and offering opportunities for organisational development. The focus of the learning and development opportunities (both professional and organisational) were specified as relating to assessment, moderation, and teaching and learning.

5.3.1.5 Educational quality

Internal moderation was also seen to function *in-practice* on educational quality. This function focuses on the students' education and qualifications layer of the contextual framework (Figure 5.3). It is a broad *in-practice* function of internal moderation as it addresses the quality of teaching and student learning aspects of the teaching, student learning, and assessment cycle. The strength of the correlations between the educational quality internal moderation scales (see below), and with *Organisational quality assurance*, *Providing professional learning opportunities*, and *Assisting organisational development,* and with *Maintaining public confidence* for PTEs (4.1.4.1, Table 4.10), suggests that academic leaders tend to consider the *in-practice* function pertaining to educational quality to be part of the broader group of purposes, more than in terms of aligning with either accountability or improvement. Educational quality featured in both data sets.

Checking educational quality and *Improving educational quality* were factors extracted from the quantitative data (see 4.1.1). The items associated with both factors were all relatively strongly supported (see Figures 4.8 and 4.9), although there was more variation in respondents' support for those items associated with improving educational quality than with most other internal moderation items. *Checking educational quality* and *Improving educational quality* were found to be strongly and positively correlated (4.1.4.1). Although extracted as two factors, the strength of the correlation suggests that academic leaders tend to see the functions of 'checking' and 'improving' as parts of a bigger, more encompassing, and holistic purpose: that of maintaining educational quality, much as perceptions of the *in-principle* function (5.2.2).

Educational quality also featured in the qualitative data (see 4.4.5.2). Respondents saw internal moderation as having a role in aspects of educational quality relating to the programme, curriculum (including taught curriculum), and teaching. They also saw internal moderation as having a role in aspects of educational quality pertaining to student learning, progress, and achievement, and the monitoring, maintenance, and assurance of these aspects.

5.3.2 NZQA moderation, in practice

Academic leaders were found to perceive NZQA moderation to function *in-practice* in four main areas: *assessment quality*, *maintaining public and stakeholder confidence*, *organisational quality assurance*, and *professional learning and development*. These areas largely reflect those that internal moderation was seen to function in (albeit with differences in scope for several). The functions occur simultaneously in all layers of the contextual framework (Figure 5.4). Nuances were found within the four areas, which are outlined as each area is discussed.
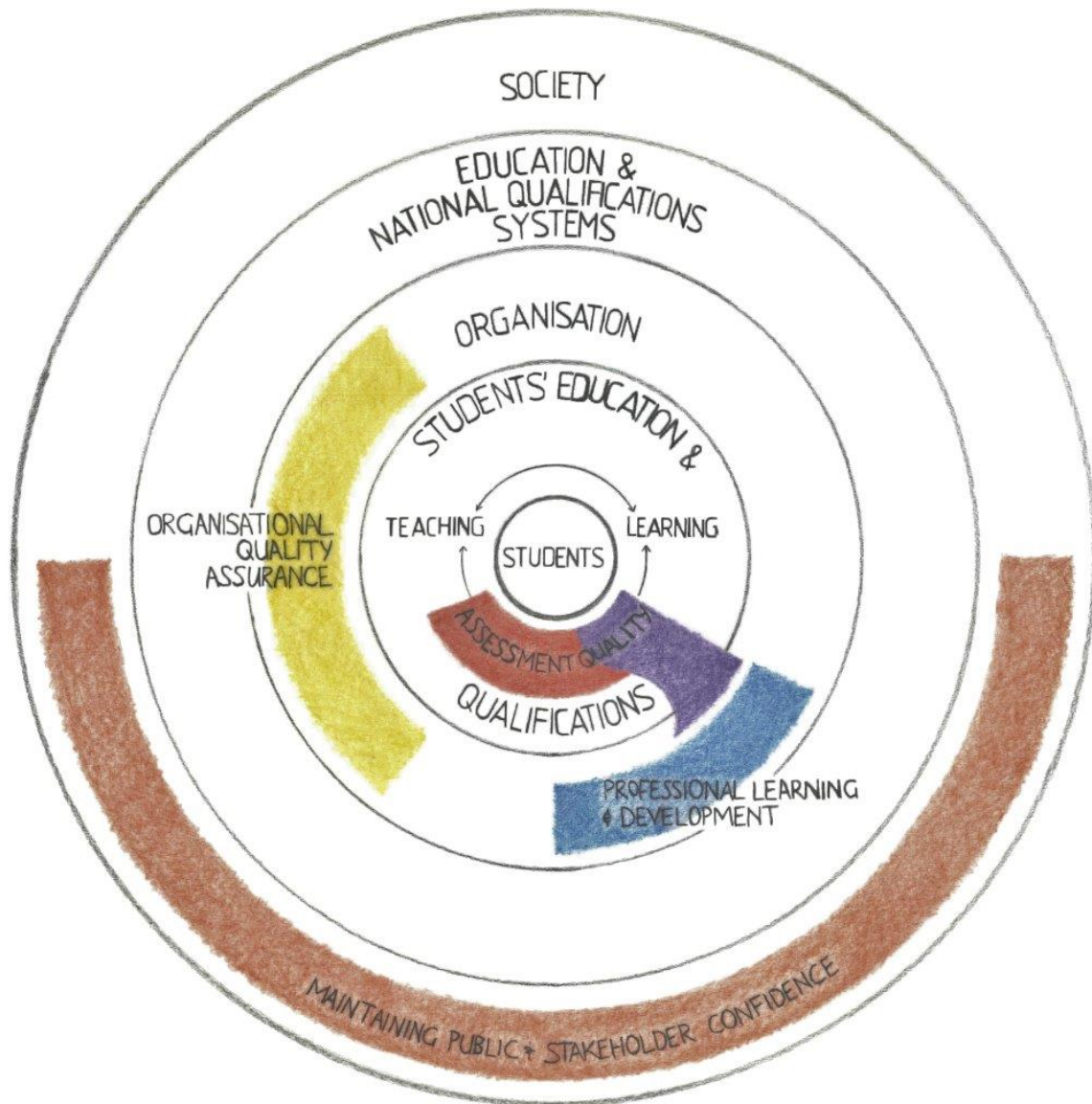
The *in-practice* functions of NZQA moderation were found to differ in breadth.[120] The narrowest function as seen by academic leaders was that of *checking* assessment quality (a function within the *assessment quality* area; see Table 4.11). Other areas are broader in focus than that of checking assessment quality, or address or act in areas that are peripheral or related to assessment quality. (See 4.2.1; Tables 4.12, 4.13, 4.14, 4.15.)

Alignment with quality control or improvement purposes (refer 2.4.3 and 2.4.4) appeared to be more pronounced in academic leaders' perceptions of the *in-practice* functions of NZQA moderation than in their perceptions of the *in-practice* functions of internal moderation or the *in-principle* functions of moderation. The narrow function of checking the quality of internal assessments clearly and cleanly aligns with a quality control purpose. The three functions of improving the quality of internal assessments, learning opportunities provided by NZQA moderation reports, and organisational quality assurance, were strongly (in two cases) and moderately (in the other) positively correlated with each other, indicating that they are associated functions (see 4.2.4.1, Table 4.16). This group of functions appears to align primarily with an improvement purpose and works in two layers of the contextual framework (the organisational layer, and the students' education and qualifications layer; Figure 5.4).

---

[120] Although to a lesser degree than for the *in-principle* functions of moderation (see 5.2) or the *in-practice* functions of internal moderation (refer 5.3.1).

Two of the functions (improving assessment quality and organisational quality assurance) have a broad focus (see Tables 4.12 and 4.14). These are discussed further below.

Figure 5.4   Perceived *in-practice* functions of NZQA moderation

5.3.2.1 Assessment quality

In practice, NZQA moderation was seen by academic leaders to directly address the quality of internal assessment. In doing so, this conception has NZQA moderation working primarily at the students' education and qualifications layer of the contextual framework (Figure 5.4), and addressing the assessment part of the teaching, student learning, and assessment

cycle. The improvement of internal assessment function (discussed below) also works at the organisation layer of the framework.

The quantitative data revealed differentiation in how academic leaders saw the role of NZQA moderation regarding assessment quality, illuminating a sharp distinction between the very narrow 'checking' (quality control) of assessment quality, and the broader function of improvement of assessment quality.[121] The correlation between the two factors, *Checking internal assessment quality* and *Improving internal assessment quality,* while statistically significant (see 4.2.4.1), was the weakest correlation found between any NZQA moderation factors, emphasising this distinction. There was almost unanimous consensus in the perception of NZQA moderation placing high emphasis on the items associated with *Checking internal assessment quality*. The perceived levels of emphasis placed on the items associated with *Improving internal assessment quality* were variable*,* which also lends weight to the inference that academic leaders see these two functions as distinct (refer 4.2.1; Figures 4.11 and 4.12).

The correlations between *Checking internal assessment quality* and almost all other NZQA moderation factors and all internal moderation factors were weak (see 4.2.4.1, 4.3.1; Tables 4.16 and 4.17). The lack of strength of these correlations suggests that, in practice, academic leaders see the quality control of internal assessment by NZQA moderation as separate and essentially a 'stand-alone' function. In other words, academic leaders' perceptions of this function do not appear to be influenced by their perceptions of the other functions of NZQA moderation or of internal moderation. The checking of internal assessment quality is a narrow function of NZQA moderation, focusing on assessment materials and assessor judgements doing the bare minimum against the requirements of each standard (see 4.2.1, Table 4.11). In contrast, the function of improving assessment quality is somewhat broader, encompassing not only the assessment materials and judgements, but also aspects of professional learning and development among personnel, and whether an assessment supports teaching and student learning (see 4.2.1, Table 4.12). Thus, the function of checking assessment quality focuses narrowly on the assessment part of the teaching, student learning, and assessment cycle within the students' education and qualifications layer of the framework. The function of improving assessment quality has a broader focus that also encompasses the teaching and student learning parts of the cycle in its scope, as well as also working in the organisational layer of the contextual framework (Figure 5.4). As such, the role of NZQA moderation in improving assessment quality is less

---

[121] Although this distinction was not reinforced through the qualitative data.

restricted in scope than the internal moderation counterpart: The focus of internal moderation is seen to be firmly on the assessments within the students' education and qualifications layer of the framework, to improve assessment materials and assessor judgements (see 4.1.1, 5.3.1.1).

Assessment quality featured in a minor way in the qualitative findings for NZQA moderation; it appears that most respondents did not have anything to add beyond what was covered by the closed-response items (4.4.1.3). Of those who commented, most respondents agreed (if reluctantly) that NZQA moderation has a role *in-practice* regarding assessment quality. They referred to the component of the assessment process that NZQA moderation addresses (assessment-in-general and assessor judgements),[122] and the properties of assessment that NZQA moderation is concerned with, but not the action that NZQA moderation takes on the assessment components. The properties of assessment with which NZQA moderation is concerned, as mentioned by respondents, included consistency (in marking and assessment methodology, within and between organisations), and assessments having robust quality, being fair, and assessing what they were meant to assess at the correct level. It seems probable that academic leaders also saw NZQA moderation as being concerned with the accuracy of assessor judgements.[123]

5.3.2.2 Maintaining public and stakeholder confidence

Maintaining public and stakeholder confidence was the second area in which academic leaders saw a role for NZQA moderation in practice. This is a broad area that encompasses public and stakeholder confidence in the internal assessment that occurs within individual organisations, the education and qualifications of graduates, and national qualifications. Thus, it sits at the societal layer of the contextual framework (Figure 5.4). It was only weakly-to-moderately correlated with the other NZQA moderation functions, suggesting that it 'stands alone' to a certain extent; that is, how academic leaders consider it tends to be relatively independent of how they consider other functions of NZQA moderation (see 4.2.4.1, Table 4.16). Maintaining public and stakeholder confidence featured in both the quantitative and qualitative data sets.

---

[122] No respondents mentioned assessment materials. In the quantitative data no distinction was made between assessment materials and assessor judgements: Items relating to both were associated with the same factors (see 4.2.1).
[123] The accuracy of assessor judgements was not mentioned as a property in relation to NZQA moderation; however, it was addressed by a closed-response item (and seen to receive high emphasis in NZQA moderation by almost all respondents, see 4.2.1). As such, the assumption is made that respondents did not feel the need to comment about it.

The factor *Maintaining public confidence* was extracted from the quantitative data and there was a widely shared belief that NZQA moderation puts a lot of weight on the associated items (see 4.2.1). NZQA moderation was seen to place more emphasis on maintaining public confidence in national qualifications than the importance that respondents placed on this function for internal moderation (see 4.3.2).

The maintenance of public and stakeholder confidence as a function of NZQA moderation was also present in the qualitative data (see 4.4.2.2). Various perspectives were put forward that ranged from endorsing NZQA moderation in this role, expressing dissatisfaction with its effectiveness, to disagreeing with the notion.

5.3.2.3 Organisational quality assurance

Organisational quality assurance was an area in which NZQA moderation was also seen by academic leaders to function in practice. This is a broad area encompassing a range of aspects of organisational quality assurance,[124] including providing information for other (internal and external) quality assurance processes as well as for performance appraisals, satisfying external requirements, and maintaining the confidence of management or governance teams. It inhabits the organisation layer of the contextual framework (Figure 5.4). The strength of the correlations between the relevant quantitative factor and the rest of the NZQA moderation factors indicates that this function aligns relatively closely with a broad, improvement purpose. However, this function was also moderately positively correlated with both separate, 'stand-alone' functions: the narrowest quality control function (checking the quality of internal assessment) and the maintaining public confidence function (see 4.2.4.1, Table 4.16). Thus, it also appears to align to a certain extent with accountability and quality control.

The factor *Organisational quality assurance* was extracted from the quantitative data, and respondents indicated that NZQA moderation places considerable weight on six of the seven associated items (Figure 4.14). In contrast, the seventh associated item (which addressed the provision of information for performance appraisals) was considered to be only weakly emphasised (see 4.2.1). As noted previously, internal moderation was rated as being more important for providing performance appraisal information than the level of emphasis on doing so that NZQA moderation was seen to hold (see 4.3.2 and 5.3.1.3).

---

[124] Like the *in-practice* function of internal moderation, it is more limited in scope than the *in-principle* function of moderation of quality assurance (see 5.3.1.3).

Organisational quality assurance featured only lightly in the qualitative data; not many respondents commented on it specifically in relation to NZQA moderation. Of those who did, their comments focused on NZQA moderation as compliance with external (i.e., NZQA) requirements, both by the organisation and for NZQA. The fallibility of the NZQA moderation system also featured in the data, as did using NZQA moderation to inform self-assessment within organisation.

5.3.2.4 Professional learning and development

The final area in which NZQA moderation was seen by academic leaders to function in practice was professional learning and development. While this function occurs in the organisation layer of the contextual framework (Figure 5.4), it is relatively narrow, pertaining mainly to learning opportunities provided through moderation reports. By definition, this function is aligned with an improvement purpose, and the strength of correlations between the quantitative factor involved in this area and other NZQA moderation functions endorses this alignment (refer 4.2.4.1, Table 4.16). The factor *Providing opportunities to learn from NZQA moderation reports* was extracted from the quantitative data. NZQA moderation was thought to place relatively strong emphasis on the associated items (see 4.2.1, Figure 4.15).

Professional learning and development also featured in the qualitative data, although it was rather contentious (4.4.4.3, 4.4.4.4). Most respondents who made comments in this regard did so to reject the notion that NZQA moderation has a role in providing these opportunities. Respondents cited the poor nature and quality of feedback provided in NZQA moderation reports as their reason for doing so. However, respondents also referred to learning and development opportunities provided by NZQA outside the arena of distance moderation, for example, via workshops. While most who mentioned these other opportunities also expressed dissatisfaction with them, that the opportunities were raised suggests that respondents recognised that these opportunities were sometimes provided.

5.3.3 Summary of perceived *in-principle* and *in-practice* functions of moderation

To summarise, academic leaders perceived that the *in-principle* functions of moderation were in the areas of *assessment quality*, *educational quality*, *professional and organisational*

learning and development, *organisational and wider quality assurance,* and *maintaining public and stakeholder confidence*. The *in-practice* functions of internal moderation were seen to be in the areas of *assessment quality*, *maintaining public and stakeholder confidence*, *organisational quality assurance*, *professional and organisational learning and development*, and *educational quality*. NZQA moderation was seen to function in the areas of *assessment quality*, *maintaining public and stakeholder confidence*, *organisational quality assurance*, and *professional learning and development*, but not in educational quality (refer Table 5.1). As such, moderation *in-principle* was seen to operate at all layers of the contextual framework (Figure 5.2), while both forms of moderation *in-practice* were seen to operate at most framework layers (Figures 5.3 and 5.4). The *in-principle* function of *organisational and wider quality assurance* encompassed the *in-practice* functions of *organisational quality assurance*, but also extended beyond and into aspects of the education and national qualifications systems. In both types of moderation *in-practice*, only the function of checking the quality of internal assessment aligned cleanly and clearly with one or other of what is often purported to be the two fundamental purposes of moderation (in this case, a quality control purpose; see 2.4.3 and 2.4.4). Academic leaders appeared to view the other functions in a holistic and encompassing way, where even if there is an emphasis on improvement, each function also impacts on quality control or accountability (and vice versa). All *in-principle* functions appeared to be holistic; none appeared to align solely with accountability or improvement.

Academic leaders' perceptions of the *in-principle* functions of moderation and the *in-practice* functions of both internal and NZQA moderation appeared to differ in their breadth of focus. These ranged from being very narrow and focused on the quality of internal assessment, through to being broader and working in adjacent or peripheral areas and contexts, such as those that pertain to the organisations within which internal assessment is conducted (refer Table 5.1). The correlations between internal moderation and NZQA moderation functions (see 4.3.1, Table 4.17) reinforce the notion of moderation having both narrow and broad functions *in-practice*. The strength of correlations between the NZQA moderation function of *Organisational quality assurance,* and the internal moderation functions of *Organisational quality assurance,* the two *educational quality* factors (*checking* and *improving*), and the two professional and organisational learning and development factors (*Providing professional learning* opportunities and *Assisting organisational development*), suggest that academic leaders see these functions as operating together. They are all broad functions that focus on aspects that are peripheral to, form the context of, or impact on the quality of internal assessment, and work at two layers of the contextual framework: the organisation layer, and the students' education and qualifications layer. In contrast, the weakness of correlations

between the NZQA moderation function of *Checking internal assessment quality*, and all internal moderation functions, indicates that academic leaders see this very narrow function of NZQA moderation as separate from the functions of internal moderation and the other functions of NZQA moderation, and thus as stand-alone.

Table 5.1   Summary of *in-principle* and *in-practice* functions of moderation

	In-practice functions	
In-principle* functions	**Internal moderation**	**NZQA moderation**
Assessment quality • Moderation-in-general has a role in assessment quality *Students' education & qualifications*	Assessment quality Checking assessment quality; Improving assessment quality • Internal moderation has a role in assessment quality • Checking assessor judgement quality factor • Checking assessment material quality factor • Improving assessment material and assessor judgement quality factor *Students' education & qualifications*	Assessment quality Checking internal assessment quality • NZQA moderation has a role in assessment quality • Checking internal assessment quality factor *Students' education & qualifications* Improving assessment quality • NZQA moderation has a role in assessment quality • Improving internal assessment quality factor *Students' education & qualifications; Organisation*
Maintaining public and stakeholder confidence • Moderation-in-general has a role in maintaining public and stakeholder confidence *Society*	Maintaining public and stakeholder confidence • Internal moderation has a role in maintaining public and stakeholder confidence • Maintaining public confidence factor *Society*	Maintaining public and stakeholder confidence • NZQA moderation has a role in maintaining public and stakeholder confidence • Maintaining public confidence factor *Society*
Organisational and wider quality assurance • Moderation-in-general has a role in quality assurance *Organisation; Education & national qualifications systems*	Organisational quality assurance • Internal moderation has a role in organisational quality assurance • Organisational quality assurance factor *Organisation*	Organisational quality assurance • NZQA moderation has a role in organisational quality assurance • Organisational quality assurance factor *Organisation*
Professional and organisational learning and development • Moderation-in-general has a role in professional and organisational learning and development *Organisation*	Professional and organisational learning and development • Internal moderation has a role in professional and organisational learning and development • Providing professional learning opportunities factor • Assisting organisational development factor *Organisation*	Professional learning and development • NZQA moderation's role in providing professional learning opportunities • Other professional learning and development opportunities provided by NZQA • Providing opportunities to learn from NZQA moderation reports factor *Organisation*
Educational quality • Moderation-in-general has a role in maintaining educational quality *Students' education & qualifications*	Educational quality • Internal moderation has a role in maintaining educational quality • Improving educational quality factor • Checking educational quality factor *Students' education & qualifications*	

Notes: Results shown as bulleted; Contextual layer(s) indicated in italics

5.4 Extent of differences observable between organisation types regarding academic leaders' perceptions of *in-practice* functions of moderation

5.4.1 Internal moderation

Differences between the ratings of academic leaders from different organisation types were evident for some functions of internal moderation but not others (see 4.1.3; shown in Table 5.2). For the internal moderation functions of *Organisational quality assurance*, *Checking educational quality*, and *Improving educational quality,* PTEs were found to rate the associated items as having higher importance than the other two organisation types. PTEs rated the *Organisational quality assurance* and *Checking educational quality* items as having significantly higher importance than either ITPs or schools. PTEs also rated the *Improving educational quality* items as having significantly higher importance than schools, but not significantly higher than ITPs.[125] These differences in perception regarding the importance of the role of internal moderation in educational quality were also suggested by the prevalence and tenor of comments in the qualitative data (see 4.4.5.4), but not its role in organisational quality assurance (4.4.3.4).

Furthermore, academic leaders from PTEs tended to see internal moderation as having a broader focus than academic leaders from the other organisations. For example, for PTEs, the internal moderation function of *Maintaining public confidence* had substantially stronger correlations with all other internal moderation functions than that function had for schools or ITPs, and likewise for the correlations between *Checking educational quality, Providing professional learning opportunities,* and *Assisting organisational development* (refer 4.1.4.1, Table 4.10). The magnitude and significance of these correlations (and others) support the inference that academic leaders in PTEs tend to see internal moderation as having wider uses in practice and more of an improvement, learning, and developmental focus, than those in the other organisation types do.

---

[125] There was also no significant difference in how ITPs and schools rated these last items.

5.4.2 NZQA moderation

Differences in the perceptions of academic leaders between the organisation types were evident regarding the *Organisational quality assurance* function of NZQA moderation, but not for any of the other functions of NZQA moderation (see 4.2.3; shown in Table 5.2). Academic leaders from PTEs rated the emphasis placed by NZQA moderation on the items associated with this function as significantly higher than respondents from either ITPs or schools. However, there was no observable difference in how respondents from ITPs and schools rated those items.

Academic leaders from PTEs also appeared to see NZQA moderation as having a broader set of functions than those from other organisations and that it tended to hold more of an improvement and development focus. The magnitude and significance of several correlations between NZQA moderation functions lend weight to these impressions, for example, between *Providing opportunities to learn from NZQA moderation reports* and *Improving internal assessment quality,* and between *Maintaining public confidence, Improving assessment quality,* and *Organisational quality assurance* (see 4.2.4.1, Table 4.16).

Table 5.2   Summary of observable differences between organisation types in perceptions of *in-practice* functions of moderation.

	Internal moderation		NZQA moderation	
	Quantitative scale	**Organisation type**	**Quantitative scale**	**Organisation type**
Observable difference	Organisational quality assurance	**_PTEs_** *ITPs; Schools*	Organisational quality assurance	**_PTEs_** *ITPs; Schools*
	Checking educational quality	**_PTEs_** *ITPs; Schools*		
	Improving educational quality	**_PTEs_** *Schools*		
No observable difference	Improving assessment material and assessor judgement quality		Providing opportunities to learn from NZQA moderation reports	
	Maintaining public confidence		Checking internal assessment quality	
	Providing professional learning opportunities		Improving internal assessment quality	
	Assisting organisational development		Maintaining public confidence	

Note: Respondents from organisation type shown in boldface and underlined were found to rate the function as having significantly higher importance (internal moderation) or emphasis (NZQA moderation) than those from the other organisation(s) shown.

5.4.3 Summary of differences in academic leaders' perceptions of *in-practice* functions of moderation

In summary, the findings of the present study suggest that academic leaders from PTEs tend to have a broader view of the *in-practice* functions of internal and NZQA moderation than academic leaders from other organisation types. All aspects for which there were observable differences were related to those broader functions (educational quality and organisational quality assurance), and in all cases, PTEs rated moderation as having higher importance or emphasis than schools[126] (see 4.1.3, 4.2.3; Table 5.2). Furthermore, PTEs were found to have stronger correlations between various moderation functions—including maintaining public confidence—than the other organisation types (see 4.1.4.1, 4.2.4.1, 4.3.1). These findings, evincing a broader view of moderation in PTEs, may be reflective of the business realities facing PTEs. Many PTEs are businesses in a competitive environment and may need to use whatever information sources available to ascertain whether (and ensure that) they meet the needs of their stakeholders.

---

[126] And, in most cases, than ITPs.

6. Conclusions

This final chapter takes the findings of the present study and considers what they mean in practical terms. Initially, the goals and research design of the study are recapped, and the main contributions of the study identified. The main implications of the present findings are considered separately for organisations and quality assurance and regulatory bodies. Bounds within which these findings should be interpreted are acknowledged, and topics for further research are suggested. The penultimate section offers organisations and NZQA proposed policy and practice responses to the present findings. The thesis concludes with final remarks from the researcher.

6.1 Goals and overview of study

The study sought to explore what academic leaders in ITPs, PTEs, and state and state-integrated secondary schools in New Zealand, who are responsible for the management of moderation, perceived the functions of moderation to be in principle, and what they perceived the functions of internal moderation and NZQA moderation to be in practice. The study also sought to ascertain the extent to which there are observable differences in the perceptions of academic leaders from those three organisation types about the *in-practice* functions of internal moderation and NZQA moderation.

A pragmatic approach was taken to achieve these aims. A mixed methods sequential study design with parallel phases of data analysis was implemented. Initially a purposive sample of academic leaders, two from each of the three organisation types, were interviewed to explore their perceptions of the functions of internal moderation and NZQA moderation. A thematic analysis of the interview data informed the development of a bespoke online survey instrument, comprising open-field and closed-response items. The survey was administered using a census approach to the target populations. Qualitative and quantitative analyses were conducted on the survey data, the findings of which were used to address the research questions of this study.

6.2 Main study contributions

Through this study, a conceptual framework was developed that visually represents the layers of contexts in which moderation is perceived to function (see 5.1; Figure 5.1; Appendix 17). This contextual framework for moderation could inform thinking about and designing moderation practices and systems. The study identified that academic leaders in New Zealand perceive the *in-principle* and *in-practice* functions of moderation to extend across the contextual framework, and to relate to the quality of internal assessment, and adjacent, contextual, and peripheral (i.e., broader) areas (Figures 5.2–5.4). The study has provided evidence that academic leaders see moderation to have both narrow and broad functions and that they perceive *in-principle* and *in-practice* moderation functions similarly (5.2, 5.3).

This study has found that academic leaders see moderation both in a more encompassing way and with more nuance than as a scission of control versus improvement (as observed by some scholars, e.g., Ratcliff, 2003; 2.4.5). Such dichotomous thinking was found primarily in relation to NZQA moderation having a control function on the quality of internal assessment (4.2.1, 4.2.4.1, 5.3.2.1; Figure 5.4), and, to a lesser degree, internal moderation having control and improvement functions on assessment quality (4.1.1, 4.1.4.1, 5.3.1.1; Figure 5.3). For the rest of the *in-practice* functions and all of the *in-principle* functions of moderation, academic leaders appeared to think about moderation in a way that was holistic and encompassing, with little distinction between improvement and control (5.2, 5.3.1-5.3.3).

Academic leaders were found to perceive there to be multiple *in-principle* functions of moderation. They conceptualise the *in-principle* functions of moderation as simultaneously narrow (focusing directly on *assessment quality*; 5.2.1) and broad (working in areas surrounding, or adjacent, related, or peripheral to internal assessment; 5.2.2–5.2.5). The broad areas in which academic leaders see moderation functioning *in-principle* in are *educational quality*, *professional and organisational learning and development*, *organisational and wider quality assurance*, and *the maintenance of public and stakeholder confidence.* These broad areas are represented across the contextual framework; each layer features one or more of them (Figure 5.2). Thus, academic leaders were found to consider moderation, *in-principle*, to be a lens through which different contextual layers, structures, and processes pertaining to education and qualifications, and that surround internal assessment, can be viewed.

The *in-practice* functions ascribed to internal moderation by academic leaders closely reflect those that they identified *in-principle*: In practice, internal moderation was seen to pertain to *assessment quality* (a narrow function), and *educational quality*, *professional and organisational learning and development*, *organisational quality assurance*, and *the maintenance of public and stakeholder confidence* (broad functions). The *in-practice* function of *organisational quality assurance* is a subset of the *in-principle* function of *organisational and wider quality assurance*; the latter also encompasses aspects that sit within the education and national qualifications systems context where the former does not. The *in-practice* functions of internal moderation are represented on three of the four layers of the contextual framework (5.3.1; Figure 5.3), with no functions represented in the education and national qualifications systems layer.

The study found that academic leaders perceive that NZQA moderation has, in practice, a narrow function of checking the quality of internal assessment, which aligns with an accountability/control focus. They also see NZQA moderation to have broad functions pertaining to *improving assessment quality*, *professional learning and development*, *organisational quality assurance*, and *the maintenance of public and stakeholder confidence* (5.3.2). The first two of these broad functions have an improvement-focus, particularly in relation to internal assessment. Thus, the *in-practice* functions of NZQA moderation touch on the same three layers of the contextual framework as those of internal moderation (Figure 5.4), although with little attention paid to teaching and learning aspects of students' education. There was also a certain amount of dissent among respondents regarding the effectiveness of NZQA moderation *in-practice* for the broad functions (4.2.1, 5.3.2.1–5.3.2.4). Thus, the notable differences between the *in-practice* functions of NZQA moderation as perceived by academic leaders, and the functions perceived of moderation *in-principle* or internal moderation *in-practice*, are that the former is not seen to have an educational quality function (see 4.4.5), and that the professional learning opportunities it provides are more restricted (refer 4.4.4, 5.3.2.4).

This study also sought to fill another gap in the literature by ascertaining whether there were any differences in New Zealand academic leaders' perceptions of the functions of moderation, according to organisation type. It was found that for most *in-practice* functions, there were no observable differences between organisations. However, academic leaders from PTEs rated internal moderation as having higher importance for *educational quality* than those from schools, and higher importance for *organisational quality assurance* than those from ITPs or schools (4.1.3, 5.4.1). PTE leaders also surmised that NZQA moderation places greater emphasis on *organisational quality assurance* than ITP or school leaders do

(4.2.3, 5.4.2). Further, PTE leaders tended to hold a broader view of the *in-practice* functions of both types of moderation as compared to those from the other organisation types (5.4).

The present study sought to explore academic leaders' *perceptions* of the functions of moderation, not what *actually happens in practice.* That is, the research design drew on perception data only, and furthermore, did not ask respondents about the moderation practices enacted within their organisations. However, the differences in ratings of importance ascribed to, or perceived emphasis placed on, the various functions (4.1, 4.2) and in qualitative data (4.4) suggest that the uses made of moderation, and the moderation functions in practice, vary across the education sector, both within and between organisation types. Work on policy enactment (e.g., by Braun et al., 2010) supports the likelihood of such variation in use and function, as does the work on the role of policy actors' current understandings (e.g., by Spillane et al., 2002; 2.5.4.1).

The contextual framework for moderation that has been developed through this study (refer Figure 5.1; Appendix 17) could be used to inform thinking about moderation, and the design and evaluation of moderation systems and practices, both in New Zealand and in other jurisdictions. For those in organisations, the framework could be used as an analytical tool to assist them to identify and examine their current understandings, practices, and uses of moderation (discussed further in 6.3.1). It could be used in the evaluation of current practices and to inform the design of new systems. The framework could be used for professional learning purposes, for example, to provide a visual representation to communicate an organisation's intended uses of moderation, or as a structure through which personnel (e.g., teachers) could identify and develop their own perceptions of moderation functions. Furthermore, the contextual framework could be used by those in regulatory or quality assurance bodies to inform thinking about moderation, and the evaluation and design (or redesign) of moderation practices or systems. This includes using the framework as an analytical tool to assist with the identification and examination of the functions of moderation from their own perspective.

6.3 Implications

The policy enactment process in organisations is shaped by multiple factors, including who takes the narrator role with respect to policy work (see 2.5, 2.5.4). The interpretations of the narrators are filtered and directed by their own current understandings and frames of reference (Ball et al., 2011a; M. Maguire et al., 2015; Spillane et al., 2002). The present

study has identified the functions of internal and NZQA moderation, as perceived by the academic leaders who are likely to be in moderation policy narrator roles in New Zealand organisations. These findings have implications for organisations on one hand, and for quality assurance bodies and policy makers on the other.

This study was set in specific New Zealand education organisations, with a focus on the perceptions of certain stakeholders. People's perceptions are situated and influenced by a myriad of contextual factors (Spillane et al., 2002). As such, the findings of this study are situated in the New Zealand contexts and populations from which the study samples were drawn. Academic leaders in other jurisdictions and different education systems, or other education sectors or organisation types, may hold different perceptions from those found in this study. Thus, some implications of these findings are primarily pertinent to the organisations within the study populations in New Zealand, or to NZQA as a quality assurance body. However, just as there are commonalities in assessment and its quality assurance (including moderation) in formal education across jurisdictions and systems (refer Chapter 2), some implications of the findings of the present study may also be relevant for organisations and quality assurance bodies in other jurisdictions, and sectors or organisation types outside of the study population. Therefore, these are tentatively offered as being applicable across systems.

6.3.1 Implications for organisations

6.3.1.1 Current perceptions and enactment

The findings of this study and contextual framework developed (Figures 5.1–5.4; Appendix 17) may assist academic leaders to identify, examine, and articulate their own perceptions about: (a) the *in-principle* functions of moderation, (b) the *in-practice* functions of internal moderation, (c) the *in-practice* functions of NZQA moderation, and (d) the *in-practice* functions of other national (e.g., jurisdiction-wide) moderation. Doing so may assist them to recognise the influence that these perceptions have on policy enactment within their organisations, and the extent to which these perceptions are evinced in their organisation's moderation practice. Further, engaging in this critical reflection may enable academic leaders to identify opportunities to modify moderation practice, or make better use of the opportunities provided by internal, NZQA, or other national moderation in their organisations.

According to Ball et al. (2011a), the perceptions of those in policy roles other than that of narrators also mediate policy enactment. Thus, the contextual framework of moderation (Figure 5.1; Appendix 17) may provide a structure through which to explore those other policy actors' perceptions of the functions of moderation. Doing so could also provide insights into each organisation's policy enactment process. An example of a practical activity is that different staff members could map their individual perceptions onto the framework, before comparing their perspectives with each other. This mapping exercise could serve to provide a point of discussion to enlarge understandings of the functions of moderation, and ways in which such functions could be enacted.

The framework and findings may aid academic leaders in the critical reading of policies set by regulatory or quality assurance bodies, identification and interrogation of core policy messages, and consideration of how these messages compare to academic leaders' current or previous interpretations. This critical reading and interrogation process may assist academic leaders to formulate how best to interpret, translate, and enact policies.

The framework and findings may also provide insights that could inform the development and implementation of moderation systems and practices within organisations that do not already have such systems in place, both in New Zealand and in other jurisdictions.

6.3.1.2 Academic leaders who currently hold a narrow view of the functions of moderation

If an academic leader currently holds a narrow view of the functions of moderation (e.g., that it acts solely in quality control of internal assessment), this contextual framework (Figure 5.1) and the visual representations of the present findings (Figures 5.2–5.4) may assist them to recognise:

a) that moderation can provide a lens through which to view broader, adjacent, related, or peripheral areas to assessment. In doing so, moderation can inform different organisational processes and offer insights into things other than internal assessment *per se*. Thus, there may be valuable opportunities for learning and development, and for quality control, that are currently being missed;

b) that moderation can have *both* quality control *and* improvement purposes, and that these purposes are not mutually exclusive;

c) that summative assessment generally does not occur in isolation (see 2.1.1). It is part of the teaching, student learning, and assessment cycle. Summative assessment is associated with, or seated within, programmes of learning. Furthermore, students will take learning from assessment experiences into their next learning (and assessment) experiences (Broadfoot, 2007; Miller et al., 2009). Teachers and assessors will also take learning with them from each experience (see 2.4.4.2; Crimmins et al., 2016; O'Connell et al., 2016).

If a narrow view of moderation (as acting solely for control of assessment quality) is reflected in an organisation's enactment and use of moderation, there may be opportunities currently being missed for learning, development, improvement, or quality control. Previous findings from literature indicate that to enable internal moderation to fulfil different functions there may be implications for the approach, scheduling, and timing of moderation, the assessment components moderated, and moderation feedback (refer 2.4).

6.3.1.3 Academic leaders who currently hold a broad view of the functions of moderation

The present findings also have implications for academic leaders who currently hold a broad view of the functions of moderation. If a broad view is reflected in the enactment and uses of moderation within an organisation, there may be a risk that moderation is 'spread so thinly' that it compromises the fulfilment of all of its functions. The framework and findings may assist academic leaders to recognise that moderation can afford an *indirect* view of aspects adjacent, related, and peripheral to summative internal assessment. However, these areas are not the prime focus of the moderation lens; instead, they should be specifically addressed by other quality assurance processes. Thus, the present findings may serve to remind those with a broad view of the functions of moderation to maintain sufficient focus on the quality of internal assessment through moderation.

6.3.1.4 Opportunities for professional learning and development

The present findings allude to the importance of organisations accessing the resources and information provided by NZQA,[127] TKI,[128] and other authoritative sources (e.g., in other jurisdictions) that pertain to standards. These resources can assist in clarifying the intention or requirements of a standard and its associated assessment process. The findings also connote the importance of having staff attend workshops and training offered by NZQA (or other authoritative bodies), and other relevant professional support. It is important that those who attend training then bring back and disseminate their learnings in their own organisations and communities of practice (as discussed by Crisp, 2017, and Grant, 2012; 2.4.4.2.10). Thus, the study highlights the importance of organisations taking whatever opportunities are available to learn from NZQA and other national moderation, and from NZQA workshops and other professional support bodies.

6.3.2 Implications for quality assurance and regulatory bodies

6.3.2.1 Policy perspective

The present study provides insight into the understandings of academic leaders in New Zealand of the functions of moderation (5.2, 5.3), which, in turn, provide insight into moderation policy enactment. The differences in academic leaders' perceptions from different organisation types found (5.4), suggest that there may be some variation in moderation policy enactment across the organisation types.

The identification of the functions that academic leaders believe moderation to have (5.2, 5.3) enables comparison with existing policy to ascertain the level of alignment. Doing so may provide insights into current policy enactment in New Zealand, especially divergence from policy intentions.

The illumination of aspects of academic leaders' current understandings and perceptions in the present study may inform policy development and revision.[129] The work of Spillane et al. (2002; 2.5.4.1) suggests that people are more likely to recognise and interpret policy messages that are familiar to them. It would follow that academic leaders in New Zealand

---

[127] Accessible through the NZQA website: www.nzqa.govt.nz
[128] Accessible through the TKI website: www.tki.org.nz
[129] It may also inform the work of the teams within NZQA that conduct MNA audits and external evaluative reviews of organisations (2.3.1.2, 2.3.1.3.1).

would be more likely to recognise and 'pick up on' moderation policy messages that reflect functions of moderation identified in this study. Further, if a new or revised policy intention was for moderation to have a function that was different from those identified in the present study (i.e., an unfamiliar function), Spillane and colleagues' cognitive framework suggests that academic leaders may not recognise those policy messages. Unfamiliar policy messages are likely to be 'lost in translation' because they are outside leaders' existing frames of reference. It follows that the subsequent enactment would likely diverge from the intent. Therefore, any policy intentions that were unfamiliar to the academic leaders would need to be accompanied by strong and clear communication, support and guidance to increase the likelihood that the leaders would recognise and 'pick up on' those policy messages.

The present explication of academic leaders' perceptions of the functions of moderation may also assist quality assurance and regulatory bodies in New Zealand to determine what information to provide, and how to refine the guidance supplied, to support policy enactment in the study populations. For example, if the message pertained to a moderation function identified in this study there would be an increased likelihood that it would be picked up and interpreted by policy narrators, and that the resulting enactment would more closely reflect policy intentions than if the message had not been recognised. The differences between perceptions of academic leaders from different organisation types suggest that there are also likely to be differences in policy enactment in the different organisation types, which may guide the tailoring of messages to the different organisation types.

Further, these findings provide insights that may be useful to quality assurance and regulatory bodies in other jurisdictions that are introducing or refining moderation of internal standards-based assessment as part of a quality assurance regime. For example, the perceived functions revealed here could provide a starting point to work from when developing or revising policies, or when seeking to ascertain stakeholders' perceptions of the functions of moderation. The findings could also provide a starting point when considering ways in which the range of moderation functions can be supported in organisations, and how to tailor associated communication.

6.3.2.2 NZQA as a quality assurance body that conducts national moderation

Present findings provide important information to NZQA about the perspectives of academic leaders—who are arguably key stakeholders—regarding the functions, emphasis, and effectiveness of NZQA moderation (5.3.2; Figure 5.4). All academic leaders participating in the present study recognised the narrow, quality-control-focused function of checking the quality of internal assessment (4.2.1). The lack of dissenting comment in the qualitative data regarding NZQA moderation checking the quality of internal assessment (4.4.1.3) implies that academic leaders believe that NZQA moderation is effective in this function.[130] It is important that NZQA maintains and enhances this function (see below). It was also found that academic leaders believe that there are broad functions of NZQA moderation (e.g., organisational quality assurance, refer 5.3.2). However, the perceived emphasis that NZQA moderation places on these broad functions was mixed, both in terms of the emphasis ratings (4.2.1), and the greater volume of dissenting comments (4.4.1.3, 4.4.2.2, 4.4.3.3, 4.4.4.3, 4.4.6.1). This suggests that some academic leaders consider NZQA moderation to not be as effective for, or focused on, these broad functions, in practice, as it could be. Thus, there may be opportunities to strengthen these functions and thereby enhance the value that organisations can gain from engaging in NZQA moderation.

The findings of this study (5.2, 5.3; Figures 5.2–5.4), and the contextual framework presented (5.1; Figure 5.1; Appendix 17), may help to actualise these opportunities for broad and improvement-focused functions. The findings highlight that summative assessment occurs in context, not in isolation (see 2.1.1, 4.4.4.1, 4.4.5), and as such, internal assessment is integrally linked to teaching and student learning. The findings also allude to the professional learning involved with moderation: Teachers, assessors, and assessment designers take learning from assessment and moderation experiences and bring that learning into their future practice (refer 2.4.4.2, 4.4.4). Many academic leaders recognise the potential for professional and organisational learning and development that is afforded through NZQA moderation reports and other avenues (4.2.1, 4.4.4.3, 4.4.4.4). However, various respondents also asserted that the nature and quality of moderation feedback needs to improve to better realise this potential. Emphasising and strengthening the broad and improvement-focused functions of NZQA moderation would likely have implications for the moderation approach and practices employed by NZQA. For example, to do so might entail

---

[130] In the qualitative data, respondents tended to refer to how effective they perceive moderation to be in practice.

feedback that details suggested improvements, pertains to teaching and student learning, or commends good practice. As with any change to the nature or focus of moderation feedback, moderator training and monitoring would be required. Such an approach may necessitate an increase in the time allowed to moderate each submission, and a corresponding reduction in the number of submissions requested in the annual national moderation sample to allow this change to be resourced (or greater resourcing applied).

The environments in which NZQA and other quality assurance and regulatory bodies work contain strong accountability drivers, and public and stakeholder confidence is paramount (see 2.3, 2.3.0.2, 2.4.2, 2.4.3). The quality assurance regimes play critical roles in ensuring that the education and qualifications systems are robust and trustworthy. The present study found that compliance and quality control were of great concern to the academic leaders (4.2.1). However, there was a sense among some respondents[131] that NZQA was more concerned with using moderation to cultivate the *public perception* of quality control and accountability than with effectively assuring and maintaining assessment standards (refer 4.4.2.2).[132] A group of respondents (also mostly from PTEs) perceived shortcomings in the NZQA moderation process or requirements which, in their view, rendered the system "weak" and open to manipulation by organisations (see 4.4.6.1). Given the differences in NZQA moderation and requirements of organisations between the secondary and tertiary sectors (refer 2.4.3.2.1), sentiments of this nature are not unexpected in relation to TEOs. However, these findings signal a need for NZQA to modify its approach to TEO moderation, and requirements of TEOs regarding internal moderation, to be more effective at quality control,[133] and, therefore, to provide more robust accountability and maintenance of public and stakeholder confidence.

6.4 Issues affecting research process and results

The main limitations to this study pertained to the organisation types included in the sample, the sampling approach taken (and the resulting sample), and the data collection instrument and approach. Despite these issues, the design of the study enabled the exploration and

---

[131] Mainly from PTEs

[132] This echoes the findings of Bloxham et al. (2016) regarding appearance versus substance (see 2.4.3.3).

[133] An obvious modification would be to require TEOs to implement a random selection process for student evidence in NZQA moderation submissions, as is currently required of schools (2.4.3.2.1). However, more substantial changes are proposed in 6.6.2, to enhance the accountability function of NZQA moderation for the TEO sector.

confirmation of the perceptions of ITP, PTE, and school academic leaders of the functions of moderation in the New Zealand context. Furthermore, the sample size was sufficient to enable observations to be made regarding the extent to which there are differences in these perceptions between academic leaders in the three different organisation types.

6.4.1 Population and organisation types included in the study

The study sample included only academic leaders with responsibility for moderation in ITPs, registered PTEs, and state and state-integrated secondary and area schools in New Zealand (3.3.1, 3.4.1). The sample did not extend to other populations within those organisation types (e.g., teachers). Education organisations from the early childhood and primary sectors, and other organisation-types from the secondary or tertiary sectors (e.g., kura kaupapa Māori, universities) were not included. Hence, inferences cannot be drawn about how the findings of the present study apply to populations, sectors, or organisation types beyond those included in the sample.

6.4.2 Sampling approach taken to the survey sample

All ITPs, registered PTEs, and relevant schools were invited to participate in the online survey (3.4.1.1). The resulting sample was neither random nor stratified: It comprised those who were interested, had time, were motivated to participate, and, in the case of schools, had received the invitation (3.4.1.1, 3.4.1.2, 3.4.1.2.1). Thus, extreme views may be more prevalent in the study data than in the rest of the target population as people with strongly held view tend to be more motivated to respond. As such, a caveat on any generalisability of the conclusions is that the sample may contain some bias (Creswell, 2012).

6.4.3 Data collection instrument

The main instrument was a self-administered online survey. The use of online surveys are characterised by low response rates and few opportunities to clarify qualitative responses (Cohen et al., 2011; Creswell, 2012; Mutch, 2013). This researcher was cognisant of these potential challenges and took every care to minimise or mitigate them (3.4.2.2; 3.4.2.3; 3.4.3; 3.4.5). These efforts were largely effective: The resulting sample sizes (3.4.1.2) were sufficient to allow the original research questions to be answered (4.1-4.3; 5.3; 5.4); the

qualitative data regarding *in-practice* functions of moderation (collected via open-field questions; 3.4.2) validated the content of the closed-response items, as the *in-practice* functions revealed were congruent with those found via quantitative data (4.4.1-4.4.5); and careful analysis of the qualitative data (3.4.5) ensured credibility of findings (4.4; 5.2; 5.3). To further explore academic leaders' perceptions of moderation functions, follow-up research using in-depth interviews (or similar, to collect rich data) would be required.

6.5 Topics for further research

This was an exploratory study, and, as such, has generated avenues for further study. The study process and literature review also highlighted adjacent topics into which research is required. These topics for further research can be grouped broadly into three main areas: moderation *in practice* (what actually happens), policy enactment processes, and perceptions of moderation functions for other types of moderation and from other populations.

6.5.1 Moderation in practice: What happens in different contexts

The present study revealed what academic leaders in New Zealand *perceive* the functions of moderation to be; however, further research is required into what the functions of moderation *actually are in practice*: What organisations and quality assurance bodies in New Zealand and in other jurisdictions use moderation for, and how effective it is.[134] More specifically, further study is required into the following:[135]

1. exploration of the functions of internal moderation and NZQA moderation *in practice* within the organisations in this study's population, and how these align with academic leaders' perceptions of those functions (as per the findings of the present study).

2. ascertainment of any differences in the functions of internal moderation and NZQA moderation *in practice* between the organisation-types in this study's population, and how these align with the differences in perceptions found by the present study.

---

[134] Such research would require the use of a range of data sources, such as documentary evidence (e.g., policies, moderation reports, meeting minutes, amended assessment or teaching materials, samples of assessed student work), observation of practice (e.g., teaching, assessment, meetings, moderation), and interviews with other parties (e.g., teachers).

[135] The literature review for the present study established that no research had been conducted into these topics in the New Zealand context (some topics) or at all (other topics; see 2.4).

3. exploration of whether moderation has an impact in practice in each of the areas that this study has found that academic leaders perceive it to (5.3), and if so, the mechanisms and processes through which this occurs.

4. examination of the impact of national moderation results and feedback on teaching or assessment practice in organisations. Drawing on the notion of policy enactment by organisations (Braun et al., 2010; 2.5), investigation into how organisations interpret, translate, and enact moderation feedback is needed. For example, if an assessment activity that is isolated from the course of learning and that is fragmentary in approach is found to be compliant through NZQA moderation, what (if any) impact on practice does this have? In a similar vein, what (if any) impact on teaching or assessment practice in organisations does moderation feedback that focuses on educationally sound, integrated, meaningful, and relevant assessment practice have?

5. investigation, via empirical research, of the impact of moderation interventions on the design and development of sound, trustworthy, and reliable assessment instruments, and the accuracy and consistency of assessor judgements, considering different types of assessment standards, and different types of moderation interventions (including internal and national moderation).

6. investigation into the effectiveness of different types of, and approaches to, moderation for the purposes of quality control, accountability, quality improvement, and learning.

7. investigation into the mechanisms and processes through which moderation and associated activities result in quality improvement and learning within organisations and the education sector.

6.5.2 Policy enactment processes

The present study drew on research which suggests that policy is enacted through processes of interpretation and translation, as opposed to being directly implemented (refer 2.5.1). Research is required in the New Zealand context to test and build on the conclusions of Braun et al. (2010), and M. Maguire et al. (2015). In particular:

8. examination of policy enactment in education organisations, and identification of the factors that mediate and shape this process.

9. comparison of policy enactment in different organisation types in the New Zealand education sector, and identification of the main contextual mediating factors involved, and their impact.

10. comparison of policy intent (on the part of the regulators) and policy enactment in the education sector, and identification of the factors that have mediated that process. That is, seeking to identify factors that enable or hinder the enactment of policy in ways that reflect the original policy intent.

6.5.3 Perceptions of moderation functions: Other moderation types and other populations

As with most social science research, the present study gathered data from a defined sample from specific populations (3.3.1, 3.4.1). Thus, further research would be required to explore the perceptions of the present sample about other types of moderation, and of other populations about moderation. The survey instrument developed and used in stage two of the present study (Appendix 7) may be appropriate for use in such explorations with relatively minimal modification,[136] although piloting the modified instrument before such use is advised. The contextual framework for moderation presented here (Figure 5.1; Appendix 17) may also be transferable for use as an analytical tool in such explorations.[137] Specifically, research is suggested to explore:

11. New Zealand academic leaders' perceptions about the functions of moderation conducted by quality assurance bodies other than NZQA (e.g., ITOs). This would provide insight into potential differences in function or approach, which may be of use to organisations and to regulatory bodies.

12. the perceptions of others in the New Zealand education sector (e.g., teachers, heads of department) about the functions of internal and NZQA moderation, and comparison with the present findings. Such people often also have roles in moderation policy enactment in their organisations and, as such, their perceptions will likely influence how they undertake that policy work (refer 2.5.4). This exploration and comparison may offer further insights into policy enactment and may be useful

---

[136] Suggested modifications include: deleting items 8.1, 29.1 and 29.3 (because they did not load cleanly onto any factors in the analyses and so were omitted from further analysis: see 3.4.4.3; 4.1.2, 4.2.2); reframing item 29.2 to ask about the perceived current emphasis on it (to mirror the rest of the NZQA moderation closed-response items); including items addressing NZQA moderation having educational quality functions (e.g., to mirror items 8.2, 8.4, 9.1-9.3, 18.1-18.4 which address internal moderation; see 4.1.1); deleting Section 4 (demographic and organisational information, as not all respondents complete it; see 3.4.1.2, 3.4.2.3); plus any changes required to match current ethical requirements.

[137] The framework would also need to be trialled to ascertain its applicability and any required modifications made before use.

for academic leaders and other policy narrators to ensure that policy enactment reflects intention.

13. the perceptions of academic leaders from other jurisdictions (in equivalent organisations to those in the population of the present study) about the functions of moderation. Doing so would enable comparison with the present findings and may offer further insights into policy enactment or uses of moderation.

14. the perceptions of academic leaders from other organisation types or sectors in New Zealand and other jurisdictions (e.g., kura kaupapa Māori, universities), about the functions of moderation. Doing so would enable comparison with the present findings and may enable alignment of experiences or understanding of assessors who work across sectors. Doing so may also offer further insights into policy enactment or uses of moderation.

15. the perceptions of NZQA staff, including moderators, about the functions of internal and NZQA moderation. Doing so would enable comparison with the present findings and may indicate the level of alignment between the perceptions of NZQA staff and academic leaders (as per this study). Revelation of differences or mismatches may offer insights into potential amendments to moderation practice, approach, or associated communication that NZQA could undertake, to better meet policy intent or allow organisations to make better use of NZQA moderation outcomes.

6.6 Proposed policy and practice responses

The following policy and practice responses to the present findings are proposed. They have been synthesised in response to critique of literature and policy, and inferences drawn from the present findings, all interpreted through the lens of this researcher's professional experience,[138] and they build on the implications already discussed (see 6.3). The main stimuli that have prompted the formulation of specific aspects of these suggestions are made explicit throughout (e.g., via section links).

The salient responses for organisations are proffered here (6.6.1), although individual organisations may already be enacting aspects of these responses.[139] The suggestions are framed for organisations within the study populations; however, they are also tentatively

---

[138] As is characteristic of pragmatism (in which this study is epistemologically founded; 3.1.2) and of constructivism (from which this study draws; 3.1.4.1).

[139] The present study did not investigate the actual moderation functions or processes that occur in practice within organisations.

offered for those in other populations or jurisdictions. Evidence of NZQA's moderation practices at the time of data collection has been considered, and the suggested responses given in 6.6.2 are recommended in relation to this.[140]

6.6.1 Within education organisations

The recommended policy responses within each organisation involve quality management systems, the functions to which moderation is put, internal moderation systems and practices, and inter-organisational moderation.

The quality management system should comprise a suite of quality assurance processes, of which moderation would be one. Moderation of internal assessment would sit alongside and complement other quality assurance processes that specifically address teaching and student learning, programmes of learning, other academic matters, and personnel development (refer 1.2.2, 2.3.1.2, 2.3.1.3). The learning and improvement potential of the various processes would be realised by using the results of those processes reflexively, to inform not only the primary focus but also other aspects to which the results are applicable.[141] For example, any learning pertaining to assessment that arises from a programme delivery evaluation would be taken forward and applied in assessments for future deliveries.

Moderation would be recognised as the quality assurance process with a primary focus on the quality of internal assessment (2.4), but also providing insight into other related, adjacent, contextual, or peripheral matters (refer 5.2, 5.3). Thus, channels would be established to communicate and use relevant moderation findings to inform these other areas (e.g., professional development plans, 4.4.4; organisational quality assurance processes, 4.4.3; teaching or curriculum matters, or student learning and progress, 4.4.5).

Internal moderation would be used for a range of functions, reflecting the present findings: quality control of internal assessment, improvement of internal assessment, and broad functions pertaining to educational quality, professional and organisational learning and

---

[140] Because national moderation as conducted by other quality assurance bodies was outside of the scope of the present study (and may differ from NZQA moderation in ways such as the nature of the assessment standards or training context), caution is advised for any application of these recommended policy responses to those other situations.

[141] As similar to how moderation findings were seen by some to inform other organisational processes (4.4.3.1, 4.4.3.2).

development, organisational quality assurance, and public and stakeholder confidence (see 4.1.1, 4.4, 5.3.1). Moderation events would be scheduled periodically and purposefully throughout a programme of learning so that all stages of the assessment process can be addressed (from design of assessment approach and instrument, to results and grades awarded; refer 2.4.0.1). Such scheduling would enable calibration of teachers' assessment expectations and understanding prior to assessment, refinement of their judgements during the marking period, and checking (quality control) of marking and grades awarded after the event.[142]

To effectively enable the range of internal moderation functions to be fulfilled (as above; 4.1.1, 4.4, 5.3.1), literature suggests that implementation of a combination of approaches is necessary (see 2.4.5.1). Social moderation approaches (via moderation meetings involving teachers, with facilitation from an assessment and moderation expert if necessary) would be used to provide opportunities for teachers to develop shared understandings, calibrate their judgement, and learn from each other (refer 2.4.4.2, 2.4.4.2.4, 2.4.4.2.7; and reflecting present findings—see 4.4.4.1). Within these meetings, conversations about assessment and moderation would not be 'siloed' from those about teaching and student learning; pedagogical conversations would be considered valuable off-shoots and integrally linked to assessment and moderation conversations (echoing the sentiment of some respondents in the present study: see 4.4.5.2; and reflecting the literature: refer 2.4.4.2.2). Expert moderation—with a senior teacher in the role of moderator—would be used for quality control purposes.[143] There would be a requirement for the moderation of assessor judgements to occur before grades were confirmed.[144] An assessment and moderation expert would work alongside teachers to redevelop or refine assessment approaches and instruments where needed, as well as to assist in the enhancement of teachers' assessment and moderation expertise, reflecting the approach taken by Black and colleagues (2010, 2011; refer 2.2.3.1) in their intervention.

Inter-organisation social moderation opportunities within subject areas or disciplines would be actively sought out and engaged in.[145] These opportunities would be facilitated to encourage collegial relationships between teachers to develop, and provide a forum for

---

[142] Much as described by Crimmins et al. (2016)—see 2.4.0.1. See also 2.4.3.4, 2.4.4.2, 2.4.4.2.3, 2.4.4.2.4.
[143] As similar to the initiative described by Crimmins et al., (2016; 2.4.0.1, 2.4.1).
[144] As there currently is for schools (see 2.4.3.2.1).
[145] The perceived value of these featured in the data: see 4.4.4.1, communities of practice. For example, *"[internal moderation provides the opportunity for sharing] of good practice … between staff at different schools when we collaborate – especially in small subject areas" (School).*

professional conversations about assessment, teaching, and student learning (which, in turn, provides opportunities for professional development; refer 2.4.4.2.4, 2.4.4.2.7). Furthermore, purposeful and structured social moderation activities would be undertaken, with the aim of calibrating teachers' assessment judgement for commonly assessed standards. The evidence of professional learning and development that these opportunities provide has been traversed (see 2.4.4.2), as has the evidence provided by O'Connell et al. (2016) of inter-organisation moderation on improving comparability between assessors (refer 2.4.3.4). Moreover, these fora may also be used to provide opportunities for inter-organisational collaboration on assessment and instructional design (as alluded to in the quote from 4.4.4.1 included in the preceding footnote).

Although outside of the scope of the present study, it is acknowledged that some teachers (in the New Zealand secondary sector in particular) feel overworked (as alluded to by several respondents in the present study—see 4.4.4.1, and as reported by Ingvarson et al., 2005; Wylie, 2013). It is recognised that the responses posited above may be more time-consuming than current practice. Therefore, ways of reducing teacher workload may be investigated, for example, reducing the number of internal assessments conducted, which could involve conducting more integrated assessments[146] or reducing the number of standards assessed.[147]

6.6.2 NZQA

The proposed policy response from NZQA involves changes to the national moderation regime (as implemented at the time of data collection: outlined in 2.4.3.2.1), and increased specificity in the requirements of TEOs in their internal moderation (as compared to those which were in place at the time of this study). These changes would strengthen the accountability and quality control functions of NZQA moderation, which may address present perceptions of NZQA moderation as lacking effectiveness in the maintenance of public and stakeholder confidence (refer 4.4.2.2) and some of the dissatisfaction with NQA moderation that was expressed (see 4.4.6.1). Such changes would also strengthen the quality improvement and broader functions for which NZQA moderation can be used by organisations, and in doing so, may alleviate the perceptions found that it had less emphasis

---

[146] Where the assessments for more than one standard are integrated.
[147] Anecdotally, many schools assess students against more standards than are required to fulfil the qualification(s) sought.

on—or effectiveness for—these functions.[148] Supporting the proposed policy responses with effective communication, guidance, and workshops (in ways described below) is also recommended to maximise the likelihood that the enactment of these proposals and response to them within education organisations reflect the policy intentions (2.5, 2.5.4.1, 6.3.2.1).

6.6.2.1 Changes to NZQA's national moderation regime

The proposed changes to NZQA's moderation regime involve modifying and extending the current two-pronged approach (see 2.4.3.2.1) in an attempt to achieve "a balance of accountability and professional learning" (Hipkins et al., 2016, p. 133), so that the first prong provides for accountability, and the second serves improvement and broader purposes.

The proffered first prong could be based on the current National Systems Check moderation (in which a representative random sample from the secondary sector is moderated annually; 2.4.3.2.1). Under this proposal, the existing National Systems Check moderation for the secondary sector would be extended to include unit standards,[149] and a parallel and equivalent regime for the tertiary sector implemented.[150] The resulting annual national agreement rates generated by secondary and tertiary National Systems Check moderation regimes would be published. This proposed prong would fulfil accountability purposes more effectively, credibly, and defensibly than the current system does.[151]

The second prong of the proposed NZQA moderation regime would involve a substantial redesign of the current school and TEO Check moderation approaches (2.4.3.2.1). Under this proposal (which has drawn from Hipkins and colleagues' recount of the early incarnation of NZQA moderation of NCEA as "a system that embodied a commitment to professional learning and providing detailed feedback to teachers" (2016, p. 129), all accountability purposes of these approaches would be stripped out. Instead, the approaches would be refocused on support, improvement, learning, and calibration of judgement (i.e., they would take on a formative purpose). Organisations would select the standards to submit (allowing

---

[148] Refer 4.2.1, 4.3.2, 4.4.1.3, 4.4.4.3, 4.4.5.3, 5.3.2.1, 5.3.2.4.
[149] From her professional experience, the researcher is aware that, at least at the time of data collection, the National Standards Check moderation only included achievement standards.
[150] It is currently limited to the secondary sector (2.4.3.2.1). See also 6.3.2.2.
[151] Implications of present findings for NZQA as a quality assurance body that conducts national moderation in terms of the strong accountability drivers present have been traversed in 6.3.2.2.

them to choose standards that they wanted feedback on),[152] and submit a selection of student samples for moderation that they considered to be around grade boundaries, or that they were unsure of (as those samples would likely generate the most useful feedback opportunities for teachers' learning; Hipkins et al., 2016). While agreement rates between moderator and assessor would be recorded, the agreement rates would not be used for monitoring purposes.[153] The proffered approach to moderation and moderation feedback for this prong (as follows) is in response to present findings (refer 4.4.4.3, 4.4.5.3, 4.4.5.4) of the rejection of the notion that NZQA moderation provides professional learning opportunities or has a role in maintaining educational quality, and attempts to ensure that this prong of NZQA moderation could be perceived by academic leaders to function in these areas, more akin to the findings for moderation-in-general (4.4.4.2, 4.4.5.1). Moderation of assessment materials would be conducted, with an additional focus on educationally meaningful and coherent assessment approaches and activities. The new organisations-specific moderation scheme would have a broader focus than solely whether the assessor judgements reflected the requirements of the standard. More fulsome moderation feedback would be given that would encompass improvement suggestions and commendations for good practice. The feedback would be designed to assist teachers to enhance their assessment practice (design, implementation, judgements), and understanding of standards, and when appropriate, their pedagogical practice or content knowledge.

Logistical issues that would need to be resolved to enable the proposed prongs to be implemented effectively[154] include the decoupling of the random selection of student samples within schools for submission for the first and second prongs.[155] Both prongs will require effective moderator training and ongoing calibration, along with robust check moderation to maximise inter-moderator consistency in judgements and feedback, within and between moderation subject systems (heeding the sentiments expressed by some respondents: 4.4.4.3, 4.4.6.1; and the recommendations of Bloxham et al., 2015; 2.4.3.3). As the proposed second prong has a different focus and approach from that which NZQA

---

[152] Parameters could be imposed to enable moderation volumes to be managed, such as specifying the number of standards from each 'moderation system' that an organisation can select.

[153] It is anticipated that the agreement rate between moderators and assessors will decrease, as moderators are more likely to disagree with grade-boundary or challenging assessor judgements than those that are straight forward (Hipkins et al., 2016). Further, present findings suggest that under the current approach, some organisations only submit samples for NZQA moderation for which they are confident of the grade awarded (see 4.4.6.1).

[154] And minimise any perverse behaviours on the part of organisations that are aimed at orchestrating an overly-positive moderation result, such as checking the marking of student work (and re-marking where necessary) before submission (2.4.3.6).

[155] For the 2015 moderation cycle onwards, the selection of assessment standards for the National Systems Check were to appear on School Check moderation plans in order that organisations could not distinguish between the two samples (NZQA, 2014g).

moderation has had for much of its history (apart from the early NCEA moderation, as above; Hipkins et al., 2016), the need for moderator training and ongoing coaching in the first number of years will be crucial. Furthermore, education of stakeholders within organisations (see 6.6.2.3, below) and NZQA[156] regarding the purposes and intentions of the different prongs would also be necessary, to ensure that the moderation results from the different prongs are used appropriately.

6.6.2.2 Increased specificity in requirements regarding internal moderation for TEOs

To address the need for changes to ensure more effective quality control (as identified in 6.3.2.2), it is proposed that NZQA impose more specific and stringent requirements on TEOs regarding their internal assessment and moderation practices than are currently in place (see 2.4.3.2.1). These more stringent requirements would include that internal moderation must occur before internal assessment results are confirmed (as is currently required of schools; see 2.4.3.2.1).

6.6.2.3 Support and guidance to the sectors

To ensure the success of these proposed changes (6.6.2.1, 6.6.2.2), it will be critical that the sectors are supported with effective communication, guidance, information, and workshops. This recommendation builds on the implications discussed in 6.3.2.1 regarding the role that familiarity with policy messages plays in the extent to which those messages are then recognised by academic leaders. The present findings suggest that the messaging and communication regarding these changes will also need to be tailored for each organisation type (refer 5.4.2, 5.4.3, 6.3.2.1). Clear and consistent explanation and differentiation of the role of each prong of the new NZQA moderation regime will be required to maximise the likelihood that academic leaders in all organisations understand the purposes of each, and interpret and use the results appropriately.[157] Informed by Grant's (2012) experience that it took time for participants to develop trust that moderation results would not be used for

---

[156] For example, the business unit that conducts the MNA audits of schools (2.3.1.2).
[157] For example, that academic leaders would, therefore, be less likely to use outcomes of the second, formative prong for inappropriate purposes, such as performance appraisals. Data pertaining to item 22.5 (*To provide information for performance appraisals*) suggest that such a use for NZQA moderation results would be familiar to at least some academic leaders (in that half of respondents rated it as having medium or high emphasis; see 4.2.1, Figure 4.1.4).

accountability purposes (see 2.4.3.6), it is anticipated that it will take some time to build the confidence of the organisations that they will not be penalised for having 'non-compliant' results in the proffered second prong of moderation.

The dissemination of focused guidance about internal moderation and other internal quality assurance processes is also advised. It seems likely that this is particularly important for some organisations in the tertiary sector as, until now, the secondary sector appears to have received more systemic support (e.g., from the Ministry of Education and TKI—refer 1.2.3— and via NZQA's MNA process; refer 2.3.1.2).[158] The present findings suggest that PTEs are more likely to view internal moderation as having broader functions than schools are (see 5.4.1). As such, PTEs are more likely to risk 'spreading moderation too thinly' (refer 6.3.1.3). Providing information and guidance regarding internal quality assurance processes that specifically address teaching and student learning matters may assist organisations to utilise internal moderation alongside the other internal quality assurance processes and use the findings of each to inform the areas to which the insights gained pertain. Conversely, the present findings suggest that schools are more likely than the other organisations to miss valuable learning and improvement opportunities by holding an overly narrow conception of internal moderation (5.4.1), and not to use the findings of moderation to inform other processes and practices (5.4.3; 6.3.1.2). Thus, guidance advising how to make broader use of moderation findings that is directed towards the school sector, is also suggested.

The final parts of the proposed response involve greater levels of collaboration between NZQA and other entities, to address (in some part) the dissatisfaction expressed by some with the professional learning and development opportunities provided by NZQA (see 4.4.4.4) and with support in the education sector (refer 4.4.6.3). It is suggested that NZQA collaborate with teaching and learning advisors/professional supports, to deliver workshops for teachers that recognise teaching, student learning, and assessment as being intrinsically related (see 2.1), and that are holistic and integrated (much as Black et al., 2010, aimed for in their intervention; 2.2.3.1). Such workshops may enable participants to recognise that moderation may be able to contribute to the maintenance of educational quality (see 4.4.5). Furthermore, it is recommended that NZQA, in conjunction with the Ministry of Education and other entities (e.g., Ako Aotearoa[159]), facilitate regional clusters of teachers from both sectors, modelled on the Scottish intervention as reported on by Grant (2012; 2.4.4.1). Such

---

[158] The prevalence of PTE respondents among those who expressed dissatisfaction with NZQA, and the nature of their complaints (refer 4.4.6), also lends weight to this inference.
[159] Ako Aotearoa (National Centre for Tertiary Teaching Excellence) provides professional development, resources, and funding for research aimed at enhancing the quality of teaching and learning in the tertiary sector in New Zealand (Ako Aotearoa, n.d.).

clusters could provide a forum for social moderation and sharing of teaching practices, and the opportunity for cross-sectoral communities of practice to develop.

6.7 Final comments

Fundamentally, the purpose of moderation is to quality-assure summative internal assessment as being robust, trustworthy, and comparable with assessments conducted in other organisations (Broadfoot, 2007; Harlen, 2007; Newton & Shaw, 2014). In doing so, moderation allows the public and stakeholders to have confidence in the qualifications to which assessments contribute, and the education of graduates involved. As is the case with most policies when enacted, multiple factors challenge the realisation of this purpose, such as other regulatory policies, funding mechanisms that exert pressure on organisations, and competing demands within organisations (Braun et al., 2011; Dalby & Noyes, 2018). The policy enactment process is also influenced by various mediating factors that exist in the organisation, including the organisational culture, resources, and the current understandings and perspectives held by those in the policy narrator roles (Ball et al., 2011a; Braun et al., 2011; Hardy, 2015; M. Maguire et al., 2015; Spillane et al., 2002; 2.5).

The present exploratory study has identified what academic leaders in New Zealand organisations (who are likely to be in policy narrator roles) believe the functions of moderation to be, both in principle and in practice. It has illuminated that academic leaders consider moderation to have narrow functions pertaining the quality of internal assessment, and broad functions pertaining to professional and organisational learning, organisational quality assurance, maintenance of public and stakeholder confidence, and in the case of internal moderation, educational quality (5.2; 5.3). The study has also revealed differences in the level of importance placed, or perceived emphasis, on some of these functions (but not on others) by academic leaders in different organisation types (5.4). Thus, the present findings provide a more comprehensive and nuanced understanding of New Zealand academic leaders' perceptions of the functions of moderation than previously existed.

These findings bring into focus the contextual layers within and on which academic leaders believe moderation works. A contextual framework for moderation has been developed (refer Figure 5.1; Appendix 17) to visually represent these different layers and assist description of the various functions of moderation in those terms (as per Figures 5.2–5.4). Thus, the contextual framework offered here provides an analytical tool through which to interpret the functions of moderation.

The findings presented in this thesis could assist regulatory bodies to design moderation policy, and tailor associated communication to the different organisation types. The findings could also inform NZQA moderation practice. Furthermore, the findings could contribute to quality assurance in organisations, by informing moderation policy design, interpretation and enactment, and informing moderation practice within and between organisations.

In closing, moderation is critical for ensuring that the high-stakes internal assessment occurring in the secondary and tertiary sectors is trustworthy, has integrity, and is conducted in such a way as to enhance—or at least not harm—teaching and student learning. In doing so, moderation supports the credibility and social acceptability of the organisations, the qualifications awarded, and of the education and national qualifications systems themselves.

References

Academic Quality Agency for New Zealand Universities. (2013). *Cycle 5 academic audit handbook for universities - version 1.0*. Retrieved from http://www.aqa.ac.nz/sites/all/files/AQA%20Cycle%205%20Academic%20Audit%20Handbook%20v1.pdf

Adie, L. (2012). Learning as identity and practice through involvement in online moderation. *Educational Assessment, Evaluation and Accountability*, *24*(1), 43–56.

Adie, L., Klenowski, V., & Wyatt-Smith, C. (2012). Towards an understanding of teacher judgement in the context of social moderation. *Education Review*, *64*(2), 223–240.

Adie, L., Lloyd, M., & Beutel, D. (2013). Identifying discourses of moderation in higher education. *Assessment and Evaluation in Higher Education*, *38*(8), 968–977.

Afrin, T. (2011). External and internal moderation: The other side of the story. In M. Hodis & S. Kaiser (Eds.), *Proceedings of the Symposium on Assessment and Learner Outcomes* (pp. 2–14). Wellington, N.Z.: Victoria University of Wellington.

Ako Aotearoa. (n.d.). Ako Aotearoa National Centre for Tertiary Teaching Excellence. Retrieved 5 August 2018, from https://akoaotearoa.ac.nz

Allal, L. (2013). Teachers' professional judgement in assessment: A cognitive act and a socially situated practice. *Assessment in Education: Principles, Policy & Practice*, *20*(1), 20–34.

Altrichter, H., & Kemethofer, D. (2015). Does accountability pressure through school inspections promote school improvement? *School Effectiveness and School Improvement*, *26*(1), 32–56.

Ball, S., Maguire, M., & Braun, A. (2012). *How schools do policy: Policy enactment in secondary schools*. Abingdon, Oxon, U.K.: Routledge.

Ball, S., Maguire, M., Braun, A., & Hoskins, K. (2011a). Policy actors: Doing policy work in schools. *Discourse: Studies in the Cultural Politics of Education*, *32*(4), 625–639.

Ball, S., Maguire, M., Braun, A., & Hoskins, K. (2011b). Policy subjects and policy actors in schools: Some necessary but insufficient analyses. *Discourse: Studies in the Cultural Politics of Education*, *32*(4), 611–624.

BBC. (2013, June 11). GCSEs: Gove pledges 'challenging' exam changes. Retrieved 31 July 2013, from http://www.bbc.co.uk/news/education-22841266

Beutel, D., Adie, L., & Lloyd, M. (2017). Assessment moderation in an Australian context: Processes, practices, and challenges. *Teaching in Higher Education*, *22*(1), 1–14.

Biddle, C., & Schafft, K. (2014). Axiology and anomaly in the practice of mixed methods work: Pragmatism, valuation, and the transformative paradigm. *Journal of Mixed Methods Research*, 1–15.

Biggs, J., & Tang, C. (2007). *Teaching for quality learning at university: What the student does* (3rd ed.). Maidenhead, England: Society for Research into Higher Education & Open University Press.

Bird, F., & Yucel, R. (2013). Improving marking reliability of scientific writing with the Developing Understanding of Assessment for Learning programme. *Assessment and Evaluation in Higher Education*, *38*(5), 536–553.

Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2010). Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice*, *17*(2), 215–232.

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 & Practice*, *18*(4), 451–469.

Bloxham, S., Hudson, J., den Outer, B., & Price, M. (2015). External peer review of assessment: An effective approach to verifying standards? *Higher Education Research & Development*, *34*(6), 1069–1082.

Bloxham, S., Hughes, C., & Adie, L. (2016). What's the point of moderation? A discussion of the purposes achieved through contemporary moderation practices. *Assessment and Evaluation in Higher Education*, *41*(4), 638–653.

Bogdan, R., & Biklen, S. (2007). *Qualitative research for education: An introduction to theories and methods* (5th ed). Boston, MA: Pearson/Allyn and Bacon.

Braun, A., Ball, S., Maguire, M., & Hoskins, K. (2011). Taking context seriously: Towards explaining policy enactments in the secondary school. *Discourse: Studies in the Cultural Politics of Education*, *32*(4), 585–596.

Braun, A., Maguire, M., & Ball, S. (2010). Policy enactments in the U.K. secondary school: Examining policy, practice and school positioning. *Journal of Education Policy*, *25*(4), 547–560.

Broadfoot, P. (2007). *An introduction to assessment*. New York, NY: Continuum International Publishing Group.

Brookhart, S. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, *22*(4), 5–12.

Bryant, F. (2000). Assessing the validity of measurement. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 99–146). Washington, DC: American Psychological Association.

Bryant, F., & Yarnold, P. (1995). Principal components analysis and exploratory and confirmatory factor analysis. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 99–136). Washington, DC: American Psychological Association.

Cardoso, S., Rosa, M., & Santos, C. (2013). Different academics' characteristics, different perceptions on quality assessment? *Quality Assurance in Education*, *21*(1), 96–117.

City & Guilds. (n.d.). Qualifications & apprenticeships. Retrieved 7 August 2018, from https://www.cityandguilds.com/qualifications-and-apprenticeships#fil=international

Cizek, G. J. (2016). Validating test score meaning and defending test score use: Different aims, different methods. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 212–225.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). Abingdon, Oxon, U.K.: Routledge.

Colbert, P., Wyatt-Smith, C., & Klenowski, V. (2012). A systems-level approach to building sustainable assessment cultures: Moderation, quality task design and dependability of judgement. *Policy Futures in Education*, *10*(4), 386–401.

Connolly, S., Klenowski, V., & Wyatt-Smith, C. (2012). Moderation and consistency of teacher judgement: Teachers' views. *British Educational Research Journal*, *38*(4), 593–614.

Controller and Auditor-General. (2012). *New Zealand Qualifications Authority: Assuring the consistency and quality of internal assessment for NCEA* [Performance audit report]. Wellington, N.Z.: Office of the Auditor-General.

Creswell, J. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Upper Saddle River, NJ: Pearson Education.

Crimmins, G., Nash, G., Oprescu, F., Alla, K., Brock, G., Hickson-Jamieson, B., & Noakes, C. (2016). Can a systematic assessment moderation process assure the quality and integrity of assessment practice while supporting the professional development of casual academics? *Assessment and Evaluation in Higher Education*, *41*(3), 427–441.

Crisp, V. (2012). An investigation of rater cognition in the assessment of projects. *Educational Measurement: Issues and Practice*, *31*(3), 10–20.

Crisp, V. (2013). Criteria, comparison and past experiences: How do teachers make judgements when marking coursework? *Assessment in Education: Principles, Policy & Practice*, *20*(1), 127–144.

Crisp, V. (2017). The judgement processes involved in the moderation of teacher-assessed projects. *Oxford Review of Education*, *43*(1), 19–37.

Crisp, V., & Novaković, N. (2009). Are all assessments equal? The comparability of demands of college-based assessments in a vocationally related qualification. *Research in Post-Compulsory Education*, *14*(1), 1–18.

Crooks, T. (2011). Assessment for learning in the accountability era: New Zealand. *Studies in Educational Evaluation*, *37*, 71–77.

Crooks, T., Kane, M., & Cohen, A. (1996). Threats to the valid use of assessments. *Assessment in Education: Principles, Policy & Practice*, *3*(3), 265–286.

Croxford, L., Grek, S., & Shaik, F. J. (2009). Quality assurance and evaluation (QAE) in Scotland: Promoting self-evaluation within and beyond the country. *Journal of Education Policy*, *24*(2), 179–193.

Curriculum Council. (2007). *Western Australian Certificate of Education - WACE Manual: General information for senior secondary schooling 2008*. Osborne Park, Western Australia: Author.

Cuttance, P. (2005). Quality assurance reviews as a catalyst for school improvement in Australia. In D. Hopkins (Ed.), *The practice and theory of school improvement: International handbook of educational change* (pp. 101–128). Retrieved from http://link.springer.com.helicon.vuw.ac.nz/chapter/10.1007/1-4020-4452-6_6

Dalby, D., & Noyes, A. (2018). Mathematics education policy enactment in England's Further Education colleges. *Journal of Vocational Education and Training*, 1–17.

Danø, T., & Stensaker, B. (2007). Still balancing improvement and accountability? Developments in external quality assurance in the Nordic countries 1996-2006. *Quality in Higher Education*, *13*(1), 81–93.

DeMars, C. (2010). *Item Response Theory*. New York, NY: Oxford University Press.

Dill, D. (2007). Will market competition assure academic quality? An analysis of the U.K. and U.S. experience. In D. F. Westerheijden, B. Stensaker, & M. J. Rosa (Eds.), *Quality assurance in higher education - trends in regulation, translation and transformation* (pp. 47–72). Dordrecht, The Netherlands: Springer.

Education Act 1989., Pub. L. No. 1989 No 80 (1989).

Education Review Office. (2014). *Framework for school reviews 2014*. Retrieved from http://ero.govt.nz/Review-Process/Frameworks-and-Evaluation-Indicators-for-ERO-Reviews/Framework-for-School-Reviews

Ehren, M., Altrichter, H., McNamara, G., & O'Hara, J. (2013). Impact of school inspections on improvement of schools: Describing assumptions on causal mechanisms in six

European countries. *Educational Assessment, Evaluation and Accountability*, *25*(1), 3–43.

Ehren, M., & Hatch, T. (2013). Responses of schools to accountability systems using multiple measures: The case of New York City elementary schools. *Educational Assessment, Evaluation and Accountability*, *25*(4), 341–373.

Ehren, M., Perryman, J., & Shackleton, N. (2015). Setting expectations for good education: How Dutch school inspections drive improvement. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, *26*(2), 296–327.

Ehren, M., Perryman, J., & Spours, K. (2014). *Accountability and school inspections* (No. June 2014). Retrieved from University of London, Institute of Education website: http://www.ioe.ac.uk/about/documents/About_Policies/Ehren_et_al_-_FINAL.pdf

European Association for Quality Assurance in Higher Education, European Students' Union, European University Association, European Association of Institutions in Higher Education, Education International, BUSINESSEUROPE, & European Quality Assurance Register for Higher Education. (2015). *Standards and guidelines for quality assurance in the European Higher Education Area (ESG)*. Retrieved from European Association for Quality Assurance in Higher Education website: http://www.enqa.eu/wp-content/uploads/2015/05/ESG_endorsed-with-changed-foreword.pdf

Gerritsen, J. (2017a, May 17). NZQA taking tougher line on low quality. *Radio New Zealand*. Retrieved from http://www.radionz.co.nz/news/national/330975/nzqa-taking-tougher-line-on-low-quality

Gerritsen, J. (2017b, July 26). NZQA finds 27 tertiary institutions not up to scratch. *Radio New Zealand*. Retrieved from http://www.radionz.co.nz/news/national/335834/nzqa-finds-27-tertiary-institutions-not-up-to-scratch

Gerritsen, J. (2017c, December 1). NZ's 'worst' polytech under scrutiny. *Radio New Zealand*. Retrieved from http://www.radionz.co.nz/news/national/345160/nz-s-worst-polytech-under-scrutiny

Gerritsen, J. (2017d, December 13). National standards ditched by government. *Radio New Zealand*. Retrieved from http://www.radionz.co.nz/news/national/346011/national-standards-ditched-by-government

Gilbert, C. (2012). *Towards a self-improving system: The role of school accountability.* [Thinkpiece]. Retrieved from National College for School Leadership website: http://dera.ioe.ac.uk/14919/1/towards-a-self-improving-system-school-accountability-thinkpiece%5B1%5D.pdf

Grainger, P., Adie, L., & Weir, K. (2016). Quality assurance of assessment and moderation discourses involving sessional staff. *Assessment and Evaluation in Higher Education*, *41*(4), 548–559.

Grant, L. (2012). Cross-sector moderation as a means of engaging staff with assessment and teaching pedagogy. *Policy Futures in Education*, *10*(4), 434–446.

Gustafsson, J.-E., Ehren, M., Conyngham, G., McNamara, G., Altrichter, H., & O'Hara, J. (2015). From inspection to quality: Ways in which school inspection influences change in schools. *Studies in Educational Evaluation*, *47*, 47–57.

Halliday-Wynes, S., & Misko, J. (2013). *Assessment issues in VET: Minimising the level of risk*. [Issues paper]. Adelaide, Australia: National Centre for Vocational Education Research.

Hambleton, R., Swaminathan, H., & Rogers, H. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.

Hammond, M., & Wellington, J. (2013). *Research methods: The key concepts*. Abingdon, Oxon, U.K.: Routledge.

Hardy, I. (2015). Curriculum reform as contested: An analysis of curriculum policy-enactment in Queensland, Australia. *International Journal of Educational Research*, *74*, 70–81.

Hardy, I., & Melville, W. (2018). The activation of epistemological resources in epistemic communities: District educators' professional learning as policy enactment. *Teaching and Teacher Education*, *71*, 159–167.

Harlen, W. (2007). *Assessment of learning*. London, U.K.: Sage.

Harvey, L., & Green, D. (1993). Defining quality. *Assessment and Evaluation in Higher Education*, *18*(1), 9–35.

Harvey, L., & Newton, J. (2007). Transforming quality evaluation: Moving on. In D. F. Westerheijden, B. Stensaker, & M. J. Rosa (Eds.), *Quality assurance in higher education - trends in regulation, translation and transformation* (pp. 225–245). Dordrecht, The Netherlands: Springer.

Harvey, L., & Stensaker, B. (2008). Quality culture: Understandings, boundaries and linkages. *European Journal of Education*, *43*(4), 427–442.

Hipkins, R. (2012). Assessment of naturally occurring evidence of literacy. *Assessment Matters*, *4*, 95–109.

Hipkins, R. (2013). *NCEA one decade on: Views and experiences from the 2012 NZCER National Survey of Secondary Schools*. Wellington, N.Z.: New Zealand Council for Educational Research.

Hipkins, R., Johnston, M., & Sheehan, M. (2016). *NCEA in context*. Wellington, N.Z.: New Zealand Council for Educational Research.

Hipkins, R., & Robertson, S. (2011). *Moderation and teacher learning: What can research tell us about their interrelationships?* Wellington, N.Z.: New Zealand Council for Educational Research.

Industry Training and Apprenticeships Act 1992., Pub. L. No. 1992 No 55 (1992).

Ingvarson, L., Kleinhenz, E., Beavis, A., Barwick, H., Carthy, I., & Wilkinson, J. (2005). *Secondary teacher workload study: Report*. Retrieved from Australian Council for Educational Research website: http://research.acer.edu.au/workforce/2/

Institutes of Technology and Polytechnics Quality. (2006). *Academic audit processes and guidelines* (4th ed.). Wellington, N.Z.: Author.

Johnson, B., & Christensen, L. (2008). *Educational research: Quantitative, qualitative, and mixed approaches* (3rd ed.). Thousand Oaks, CA: Sage.

Johnson, G. (2015). Record of assessment moderation practice (RAMP): Survey software as a mechanism of continuous quality improvement. *Assessment and Evaluation in Higher Education*, *40*(2), 265–278.

Johnson, S. (2013). On the reliability of high-stakes teacher assessment. *Research Papers in Education*, *28*(1), 91–105.

Kemmis, S., Cole, P., & Suggett, D. (1998). Orientations to curriculum. In E. Hatton (Ed.), *Understanding teaching* (2nd ed., pp. 139–144). Sydney, Australia: Harcourt Brace.

Kleijnen, J., Dolmans, D., Willems, J., & van Hout, H. (2013). Teachers' conceptions of quality and organisational values in higher education: Compliance or enhancement? *Assessment and Evaluation in Higher Education*, *38*(2), 152–166.

Kleijnen, J., Dolmans, D., Willems, J., & van Hout, H. (2014). Effective quality management requires a systematic approach and a flexible organisational culture - a qualitative study among academic staff. *Quality in Higher Education*, *20*(1), 103–126.

Klenowski, V. (2013). Towards improving public understanding of judgement practice in standards-referenced assessment: An Australian perspective. *Oxford Review of Education*, *39*(1), 36–51.

Klenowski, V., & Wyatt-Smith, C. (2014). *Assessment for education: Standards, judgement and moderation*. London, U.K.: Sage.

Larsen, R., & Marx, M. (1986). *An introduction to mathematical statistics and its applications* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Law, D. (2010). Quality assurance in post-secondary education: Some common approaches. *Quality Assurance in Education*, *18*(1), 64–77.

Lichtman, M. (2010). *Qualitative research in education: A user's guide* (2nd ed.). Thousand Oaks, CA: Sage.

Lillis, D. (2012). Systematically evaluating the effectiveness of quality assurance programmes in leading to improvements in institutional performance. *Quality in Higher Education*, *18*(1), 59–73.

Lissitz, R., & Samuelsen, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher*, *36*(8), 437–448.

Maguire, K., & Gibbs, P. (2013). Exploring the notion of quality in quality higher education assessment in a collaborative future. *Quality in Higher Education*, *19*(1), 41–55.

Maguire, M., Braun, A., & Ball, S. (2015). 'Where you stand depends on where you sit': The social construction of policy enactments in the (English) secondary school. *Discourse: Studies in the Cultural Politics of Education*, *36*(4), 485–499.

Maguire, M., Hoskins, K., Ball, S., & Braun, A. (2011). Policy discourses in school texts. *Discourse: Studies in the Cultural Politics of Education*, *32*(4), 597–609.

Manly, B. (1986). *Multivariate statistical methods: A primer*. London, U.K.: Chapman and Hall.

McDonald, J. (2014). *Handbook of biological statistics* (3rd ed.). Retrieved from http://www.biostathandbook.com/small.html

McMillan, J. (2007). *Classroom assessment: Principles and practice for effective standards-based instruction* (4th ed). Bolton, MA: Pearson Education.

Meade, P., & Woodhouse, D. (2000). Evaluating the effectiveness of the New Zealand Academic Audit Unit: Review and outcomes. *Quality in Higher Education*, *6*(1), 19–29.

Mertens, D. (2010). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative and mixed methods* (3rd ed.). Thousand Oaks, CA: Sage.

Metro Group. (n.d.). Metro Group. Retrieved 16 November 2013, from http://www.metros.ac.nz/

Miller, D., Linn, R., & Gronlund, N. (2009). *Measurement and assessment in teaching* (10th ed). Upper Saddle River, NJ: Pearson Education.

Millward, P., & Timperley, H. (2010). Organizational learning facilitated by instructional leadership, tight coupling and boundary spanning practices. *Journal of Educational Change*, *11*(2), 139–155.

Ministry of Education. (2010, July). National Standards factsheet: Moderation. Retrieved 16 October 2017, from http://nzcurriculum.tki.org.nz/National-Standards/Key-information/Fact-sheets/Moderation

Ministry of Education. (2015). *New Zealand education system overview*. Retrieved from https://www.education.govt.nz/assets/Uploads/NZ-Education-System-Overview-publication-web-format.pdf

Ministry of Education. (2018a). Education in New Zealand. Retrieved 21 May 2018, from https://www.education.govt.nz/ministry-of-education/our-role-and-our-people/education-in-nz/

Ministry of Education. (2018b). School structures and governance. Retrieved 6 April 2018, from https://www.education.govt.nz/school/running-a-school/school-structures-and-governance/

Ministry of Education. (2018c, March 19). Ministry funding deciles. Retrieved 15 December 2018, from Education in New Zealand website: https://parents.education.govt.nz/primary-school/schooling-in-nz/ministry-funding-deciles/

Misko, J. (2015a). *Developing, approving and maintaining qualifications: Selected international approaches* (Research Report No. 2775). Adelaide, Australia: National Centre for Vocational Education Research.

Misko, J. (2015b). *Regulating and quality assuring VET: International developments* (Research Report No. 2806). Adelaide, Australia: National Centre for Vocational Education Research.

Misko, J., Halliday-Wynes, S., Stanwick, J., & Gemici, S. (2014). *Quality assessments: Practice and perspectives* [Research report]. Adelaide, Australia: National Centre for Vocational Education Research.

Moir, J. (2017, December 13). Farewell to 'distracting' National Standards. *The Dominion Post*, p. 1.

Montecinos, C., Madrid, R., Fernández, M., & Ahumada, L. (2014). A goal orientation analysis of teachers' motivations to participate in the school self-assessment processes of a quality assurance system in Chile. *Educational Assessment, Evaluation and Accountability*, *26*(3), 241–261.

Morgan, D. (2014). Pragmatism as a paradigm for social research. *Qualitative Inquiry*, *20*(8), 1045–1053.

Morshead, R. (1995). *Patterns of educational practice: Theories of curriculum*. Ann Arbour, MI: The Pierian Press.

Moss, P., Girard, B., & Haniford, L. (2006). Validity in educational assessment. *Review of Research in Education*, *30*, 109–162.

Mottier Lopez, L., & Pasquini, R. (2017). Professional controversies between teachers about their summative assessment practices: A tool for building assessment capacity. *Assessment in Education: Principles, Policy & Practice*, *24*(2), 228–249.

Mutch, C. (2013). *Doing educational research: A practitioner's guide to getting started* (2nd ed.). Wellington, N.Z.: New Zealand Council for Educational Research.

New Zealand Qualifications Authority. (n.d.a). External moderation. Retrieved 14 November 2014, from http://www.nzqa.govt.nz/assets/_generated_pdfs/external-moderation-647.pdf

New Zealand Qualifications Authority. (n.d.b). How NCEA works. Retrieved 25 August 2013, from http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/understanding-ncea/how-ncea-works/

New Zealand Qualifications Authority. (n.d.c). Level 1 literacy and numeracy requirements from 2011. Retrieved 16 October 2011, from http://www.nzqa.govt.nz/about-us/our-role/legislation/nzqa-rules/secondary-schools-supporting-information/literacy-numeracy-2011/#note1

New Zealand Qualifications Authority. (n.d.d). Our role. Retrieved 14 October 2017, from

    http://www.nzqa.govt.nz/assets/_generated_pdfs/our-role-2.pdf

New Zealand Qualifications Authority. (n.d.e). Policy and guidelines for the conduct of

    External Evaluation and Review. Retrieved 26 October 2015, from

    http://www.nzqa.govt.nz/providers-partners/external-evaluation-and-review/policy-

    and-guidelines-eer/

New Zealand Qualifications Authority. (n.d.f). Standards. Retrieved 6 April 2015, from

    http://www.nzqa.govt.nz/qualifications-standards/standards/

New Zealand Qualifications Authority. (n.d.g). University Entrance. Retrieved 7 August 2018,

    from https://www.nzqa.govt.nz/qualifications-standards/awards/university-entrance/

New Zealand Qualifications Authority. (1992a). *Guidelines and criteria for accreditation to*

    offer national certificates and national diplomas. Wellington, N.Z.: Author.

New Zealand Qualifications Authority. (1992b). *Guidelines and criteria for registration of*

    Private Training Establishments. Wellington, N.Z.: Author.

New Zealand Qualifications Authority. (1996). *Learning and assessment: A guide to*

    assessment for the National Qualifications Framework (for consultation and

    comment). Wellington, N.Z.: Author.

New Zealand Qualifications Authority. (2004). New Zealand Diploma in Business. Retrieved

    24 November 2018, from Author website:

    https://web.archive.org/web/20050204070620/http://nzqa.govt.nz:80/qualifications/ter

    tqual/dipbus/index.html

New Zealand Qualifications Authority. (2009). *Using evaluation to strengthen organisational*

    self-assessment. Retrieved from http://www.nzqa.govt.nz/assets/Providers-and-

    partners/Registration-and-accreditation/Self-assessment/evaluation-for-self-

    assessment.pdf

New Zealand Qualifications Authority. (2010a). *Read texts with understanding* (version 1).

    Retrieved 1 November 2013, from

    http://www.nzqa.govt.nz/nqfdocs/units/pdf/26624.pdf

New Zealand Qualifications Authority. (2010b). *Use number to solve problems* (version 1). Retrieved 1 November 2013, from http://www.nzqa.govt.nz/nqfdocs/units/pdf/26623.pdf

New Zealand Qualifications Authority. (2011a). *Changes to external moderation for 2012* (Circular No. 2011/055). Retrieved from Author website: https://web.archive.org/web/20111228025302/http://www.nzqa.govt.nz:80/about-us/publications/newsletters-and-circulars/secqual/changes-to-external-moderation-for-2012

New Zealand Qualifications Authority. (2011b). *Guide to requirements for consent to assess for schools.* Retrieved from http://www.nzqa.govt.nz/assets/Providers-and-partners/Registration-and-accreditation/Guide-to-Req-for-Consent-to-Assess-for-Schools.pdf

New Zealand Qualifications Authority. (2011c). *Guidelines for approval of programmes of study leading to qualifications listed on the New Zealand Qualifications Framework and accreditation of tertiary education providers.* Wellington, N.Z.: Author.

New Zealand Qualifications Authority. (2012a). *An independent evaluation of NZQA's evaluative quality assurance framework - report of the panel.* Retrieved from http://www.nzqa.govt.nz/assets/Studying-in-NZ/Quality-assurance/evaluation-qa-framework-report.pdf

New Zealand Qualifications Authority. (2012b). *New literacy and numeracy requirements for NCEA level 2 and level 3* (Circular No. S2012/019). Retrieved from Author website: http://web.archive.org/web/20130209121337/http://www.nzqa.govt.nz/about-us/publications/newsletters-and-circulars/secqual/new-literacy-and-numeracy-requirements-for-ncea-l2-and-l3/

New Zealand Qualifications Authority. (2012c). *Update on the review of New Zealand Curriculum based standards, transition arrangements, consent to assess, moderation best practice workshops and NCEA level 1 literacy and numeracy requirements*

(Circular No. NZQF2012/001). Retrieved from http://www.nzqa.govt.nz/assets/About-us/Publications/Framework-circular/N12001.pdf

New Zealand Qualifications Authority. (2013a). *Guidelines for Private Training Establishment registration* (version 2.0). Retrieved from http://www.nzqa.govt.nz/assets/Providers-and-partners/Registration-and-accreditation/New-provider/guidelines-PTE-registration.pdf

New Zealand Qualifications Authority. (2013b). *Guidelines for random selection of student work for external moderation*. Retrieved from http://www.nzqa.govt.nz/assets/Providers-and-partners/Assessment-and-moderation/sec-mod-selection-guidelines.pdf

New Zealand Qualifications Authority. (2013c). *National external moderation manual for Tertiary Education Organisations (for NZQA-managed standards on the Directory of Assessment Standards) - 2013 TEO Manual*. Wellington, N.Z.: Author.

New Zealand Qualifications Authority. (2014a). Assessment (including examination) rules for schools with consent to assess 2014. Retrieved from http://www.nzqa.govt.nz/about-us/our-role/legislation/nzqa-rules/assessment-and-examination-rules-for-schools-2014/

New Zealand Qualifications Authority. (2014b). *External Evaluation and Review (EER) rules 2013*. Retrieved from http://www.nzqa.govt.nz/assets/About-us/Our-role/Rules/EER-Rules.pdf

New Zealand Qualifications Authority. (2014c). *Guidelines for assuring national consistency of graduate outcomes for qualifications listed on the New Zealand Qualifications Framework*. Retrieved from http://www.nzqa.govt.nz/assets/Providers-and-partners/Registration-and-accreditation/Consistency-of-graduate-outcomes/guidelines-consistency-of-graduate-outcomes.pdf

New Zealand Qualifications Authority. (2014d). National external moderation manual for Tertiary Education Organisations 2014. Retrieved

https://www.nzqa.govt.nz/providers-partners/assessment-and-moderation-of-standards/tertiary-moderation/manual-for-teo/

New Zealand Qualifications Authority. (2014e). *Statement of performance expectations 2014/15.* Wellington, N.Z.: Author.

New Zealand Qualifications Authority. (2014f). *The use of 'N' as a 'material' descriptor for external moderation* (Circular No. A2014/027). Retrieved from http://www.nzqa.govt.nz/about-us/publications/newsletters-and-circulars/assessment-matters/the-use-of-n-as-a-material-descriptor-for-external-moderation/

New Zealand Qualifications Authority. (2014g). *Updates to external moderation processes* (Circular No. A2014/051). Retrieved from http://www.nzqa.govt.nz/about-us/publications/newsletters-and-circulars/assessment-matters/the-use-of-n-as-a-material-descriptor-for-external-moderation/

New Zealand Qualifications Authority. (2015). *Guidelines for applying for approval of programmes leading to New Zealand qualifications at Levels 1-6 on the NZQF and the accreditation of organisations to provide approved programmes.* Retrieved from http://www.nzqa.govt.nz/assets/Providers-and-partners/Registration-and-accreditation/guidelines-approval-programme-accred.pdf

New Zealand Qualifications Authority. (2016a). *Guidelines for applying for consent to assess standards listed on the Directory of Assessment standards.* Retrieved from http://www.nzqa.govt.nz/assets/Providers-and-partners/Registration-and-accreditation/guidelines-consent-to-assess.pdf

New Zealand Qualifications Authority. (2016b). *The New Zealand Qualifications Framework.* Retrieved from http://www.nzqa.govt.nz/assets/Studying-in-NZ/New-Zealand-Qualification-Framework/requirements-nzqf.pdf

New Zealand Qualifications Authority. (2017a). *Consent to assess against standards on the Directory of Assessment Standards rules 2011* (Approved version 2). Retrieved from http://www.nzqa.govt.nz/assets/About-us/Our-role/Rules/consent-to-assess-rules-09-11.pdf

New Zealand Qualifications Authority. (2017b). *NZQF programme approval and accreditation rules 2013*. Retrieved from http://www.nzqa.govt.nz/assets/About-us/Our-role/Rules/prog-app-accred-rules-2016.pdf

New Zealand Qualifications Authority. (2018). Literacy requirements for University Entrance. Retrieved 7 August 2018, from https://www.nzqa.govt.nz/qualifications-standards/awards/university-entrance/literacy-requirements/

Newby, P. (2010). *Research methods for education*. Harlow, Essex, U.K.: Pearson Education.

Newton, P. (2007). Contextualising the comparability of examination standards. In P. Newton, J.-A. Bourd, H. Goldstein, H. Patrick, & P. Tymms (Eds.), *Techniques for monitoring the comparability of examination standards* (pp. 9–37). London, U.K.: Qualifications and Curriculum Authority.

Newton, P., & Shaw, S. (2014). *Validity in educational and psychological assessment*. ProQuest Ebook Central: Sage.

Newton, P., & Shaw, S. (2016). Disagreement over the best way to use the word 'validity' and options for reaching consensus. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 178–197.

Ng, P. (2007). Quality assurance in the Singapore education system in an era of diversity and innovation. *Educational Research for Policy and Practice*, *6*(3), 235–247.

O'Connell, B., de Lange, P., Freeman, M., Hancock, P., Abraham, A., Howieson, B., & Watty, K. (2016). Does calibration reduce variability in the assessment of accounting learning outcomes? *Assessment and Evaluation in Higher Education*, *41*(3), 331–349.

O'Neill, O. (2013). Intelligent accountability in education. *Oxford Review of Education*, *39*(1), 4–16.

Openshaw, R. (2014). Picot Report/Tomorrow's Schools. In *Dictionary of Educational History in Australia and New Zealand (DEHANZ)*. Retrieved from http://dehanz.net.au/wp-content/uploads/2014/01/Picot-Report.pdf

Queensland Curriculum and Assessment Authority. (2018, January 22). New Queensland

    Certificate of Education (QCE) and tertiary entrance systems. Retrieved 27 January

    2018, from https://www.qcaa.qld.edu.au/senior/new-snr-assessment-te

Queensland Studies Authority. (2010). *A - Z of senior moderation.* Retrieved from

    http://www.qsa.qld.edu.au/downloads/senior/snr_qa_mod_a-z.pdf

Rasooli, A., Zandi, H., & DeLuca, C. (2018). Reconceptualising classroom assessment

    fairness: A systematic meta-ethnography of assessment literature and beyond.

    Studies in Educational Evaluation, *56*, 164–181.

Ratcliff, J. (2003). Dynamic and communicative aspects of quality assurance. *Quality in*

    Higher Education, *9*(2), 117–131.

Redmond, A. (2018, December 13). Revolution in the classroom. *The Dominion Post*, pp.

    22–23.

Sadler, D. (2012). Assuring academic achievement standards: From moderation to

    calibration. *Assessment in Education: Principles, Policy & Practice*, 1–15.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores.

    Psychometric Monograph, (17).

School report. (n.d.). *Stuff.* Retrieved 14 February 2018, from

    https://www.stuff.co.nz/interactives/schoolreport/

Sheikh, I., & Bagley, C. (2018). Towards a policy social psychology: Teacher engagement

    with policy enactment and the core concept of Affective Disruption. *British*

    Educational Research Journal, *44*(1), 43–60.

Sirkin, R. (2006). *Statistics for the social sciences* (3rd ed). Thousand Oaks, CA: Sage.

Slowley, D. (2008). A contingent view of secondary school leadership in New Zealand.

    Journal of Educational Leadership, Policy and Practice, *23*(2), 31–40.

Smaill, E. (2013). Moderating New Zealand's national standards: Teacher learning and

    assessment outcomes. *Assessment in Education: Principles, Policy & Practice*,

    20(3), 250–265.

Spillane, J., Reiser, B., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research*, *72*(3), 387–431.

Statistics New Zealand. (n.d.). New Zealand in the OECD - Population. Retrieved 21 May 2018, from

http://archive.stats.govt.nz/browse_for_stats/government_finance/central_governmen t/nz-in-the-oecd/population.aspx

Statistics New Zealand. (2018). Population. Retrieved 21 May 2018, from

https://www.stats.govt.nz/topics/population

Stoten, D. (2012). Exploring quality assurance in sixth form colleges. *Quality Assurance in Education*, *20*(3), 259–273.

Sykes, J. (Ed.). (1982). *The concise Oxford dictionary of current English* (7th ed). U.K.: Oxford University Press.

Tabachnick, B., & Fidell, L. (2013). *Using multivariate statistics* (6th ed). Upper Saddle River, NJ: Pearson Education.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioural sciences*. Thousand Oaks, CA: Sage.

Tertiary Education Commission. (2008a). *Learning progressions for adult literacy*. Wellington, N.Z.: Author.

Tertiary Education Commission. (2008b). *Learning progressions for adult numeracy*. Wellington, N.Z.: Author.

Tertiary Education Commission. (2016, March 11). Literacy and Numeracy for Adults Assessment Tool. Retrieved 12 October 2018, from

https://www.tec.govt.nz/focus/our-focus/adult-literacy-numeracy/assessment-tool/

Thomas, G., Johnston, M., & Ward, J. (2014). *Alignment of literacy and numeracy measures: Research for the Tertiary Education Commission*. Retrieved from Tertiary Education Commission website:

http://www.tec.govt.nz/Documents/Reports%20and%20other%20documents/Alignme

nt-of-Literacy-and-Numeracy-Measures.pdf

Van Kemenade, E., Pupius, M., & Hardjono, T. (2008). More value to defining quality.

Quality in Higher Education, *14*(2), 175–185.

Vanhoof, J., Van Petegem, P., Verhoeven, J., & Buvens, I. (2009). Linking the policymaking

capacities of schools and the quality of school self-evaluations. *Educational*

Management, Administration & Leadership, *37*(5), 667–686.

Vaughan, K., Gardiner, B., & Eyre, J. (2012). *A transformational system for on-job*

assessment in the building and construction industries. Wellington, N.Z.: Ako

Aotearoa.

Ward, J., & Thomas, G. (2016). *National Standards: School Sample Monitoring & Evaluation*

Project, 2010-2014. Wellington, N.Z.: Ministry of Education.

Watty, K., Freeman, M., Howieson, B., Hancock, P., O'Connell, B., de Lange, P., &

Abraham, A. (2014). Social moderation, assessment and assuring standards for

accounting graduates. *Assessment and Evaluation in Higher Education*, *39*(4), 461–

478.

Wong, M., & Li, H. (2010). From external inspection to self-evaluation: A study of quality

assurance in Hong Kong kindergartens. *Early Education and Development*, *21*(2),

205–233.

Wyatt-Smith, C., & Klenowski, V. (2013). Explicit, latent and meta-criteria: Types of criteria at

play in professional judgement practice. *Assessment in Education: Principles, Policy*

& Practice, *20*(1), 35–52.

Wyatt-Smith, C., 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.

Wylie, C. (2013). *Secondary schools in 2012: Main findings from the NZCER national*

survey. Wellington, N.Z.: New Zealand Council for Educational Research.

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

Wylie, C., & Bonne, L. (2016). *Secondary schools in 2015: Findings from the NZCER national survey*. Wellington, N.Z.: New Zealand Council for Educational Research.

Appendices

1. Interview consent form

Academic leaders' perceptions of the functions of moderation: A mixed methods study

Consent to Participate in Research: Initial Interview

I have been given and have understood an explanation of this research project. I have had an opportunity to ask questions and have them answered to my satisfaction. I understand that I may withdraw myself (or any information I have provided) from this project before 28 February 2014 without having to give reasons.

I understand that any information I provide will be kept confidential to the researcher, the supervisors and the person who transcribes the digital recordings of our interview. I understand that the published results will not mention my name or the name of my education organisation, and that no opinions will be attributed to me in any way that will allow identification of me or my education organisation. I understand that the digital recording of interviews will be deleted five years after the end of the project.

I understand that I have the opportunity to check any transcripts made from the interview before data analysis occurs.

I understand that the data I provide will not be used for any purpose other than that stated in the information letter nor released to others without my written consent.

I would / would not like to receive a summary of the results of the research when it is completed. (Please indicate which option you select)

I agree to take part in this research.

Signature: .... **Date:** ....

Name of participant (printed): ....

Education organisation: ....

2. Interview information sheet

Academic leaders' perceptions of the functions of moderation: A mixed methods study

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

VICTORIA
UNIVERSITY OF WELLINGTON

Participant Information Sheet: Initial Interview

Researcher: Anna Williams, School of Education Policy and Implementation, Victoria University of Wellington

I am a Masters student in the Masters of Education at Victoria University of Wellington. As part of this degree I am undertaking a research project leading to a thesis. The purpose of the project I am undertaking is to examine the perceptions of academic leaders in education organisations about the functions of internal and national moderation.

I am inviting senior and middle managers in state secondary schools, Institutes of Technology and Polytechnics, and Private Training Enterprises, who have oversight of, and responsibility for, internal moderation and national moderation to participate in this study. These academic leaders will be asked to participate in a one-to-one face-to-face interview, which will be no longer than one hour in length. The interview will be recorded using a digital voice recorder, and will later be transcribed. Analysis of the interview transcripts will inform the development of an online survey, which will then be distributed to academic leaders in New Zealand state secondary schools, Institutes of Technology and Polytechnics, and Private Training Enterprises. Further analysis of the interview transcripts will be undertaken, and the findings incorporated into the thesis.

Should any interview participants feel the need to withdraw from the project, they may do so without question anytime before **28 February 2014**. Just let me know at the time.

Responses will form the basis of my research project and will be put into a written report on an anonymous basis. It will not be possible for you or your organisation to be identified personally. All material collected will be kept confidential. No person besides me, my supervisors Dr Michael Johnston and Dr Robin Averill, and the person who transcribes the interviews will see the interview transcripts or survey responses. The thesis will be submitted for marking to the School of Education Policy and Implementation, and deposited in the University Library. It is intended that one or more articles will be submitted for publication in scholarly journals, and one or more presentations may be made at academic or professional conferences. Interview recordings and transcripts, and survey responses, will be destroyed five years after the end of the project.

I am undertaking this research in the capacity of a private citizen; however, I am employed by the New Zealand Qualifications Authority in the Assessment and Moderation Services business unit, where I oversee the moderation of several systems of assessment standards. I am conducting this study completely outside of my role with NZQA.

The research has been approved by the Faculty of Education Human Ethics Sub-committee under delegated authority from the Victoria University Human Ethics Committee. If you have any ethical questions about this research, please contact Dr Allison Kirkman, Chair of the Human Ethics Committee, Victoria University of Wellington (Allison.Kirkman@vuw.ac.nz, phone 04 463 5676).

If you have any further questions or would like to receive further information about the project, please contact me at 027 449 0703 or Anna.Williams2@vuw.ac.nz, or either of my supervisors Dr Michael Johnston or Dr Robin Averill, at the School of Education at Victoria University – Michael at 04 463 9675 or Michael.Johnston@vuw.ac.nz, or Robin at 04 463 9714 or Robin.Averill@vuw.ac.nz.

Anna Williams

3. Interview schedule

Interview schedule: Initial interviews

Name: Education Organisation:

Date: Interview venue:

Introduction

Introduction, thanks for agreeing to participate, scene setting, etc

This **project aims** to examine what academic leaders with oversight of, and responsibility for, internal moderation and national moderation in education organisations believe the functions of moderation to be.

Confidential research, consent, withdrawal from project—as per Initial Interview Information sheet

Structure of interview: start with questions about Internal moderation (that is, moderation that occurs within education organisations), then questions about National moderation (as is conducted by the New Zealand Qualifications Authority), finish up with questions about you and your education organisation.

Do you have any questions before we get started?

Interview questions

1. What is your job title and role within this organisation?

Internal moderation

2. What do you think the functions of internal moderation of assessment are in your organisation? Clarify if needed: • Internal moderation is moderation that is	

conducted within organisations	

Prompt questions if needed:
 	
  a. Do you think there are any functions to do with quality assurance?
  b. If so, what are they?	

c. Do you think there are any functions that are to do with maintaining public confidence in the education and qualifications students receive? d. If so, what are they?	
e. Do you think there are any functions that are to do with providing professional learning opportunities to teachers/tutors and assessors? f. If so, what are they?	
3. In your view, are there any other functions of internal moderation of assessment (beyond those you have already discussed), and that might not occur in your organisation? (E.g. to do with quality assurance, maintaining public confidence, providing professional learning opportunities, or other functions?) g. If so, what are they?	

National moderation (as conducted by NZQA)

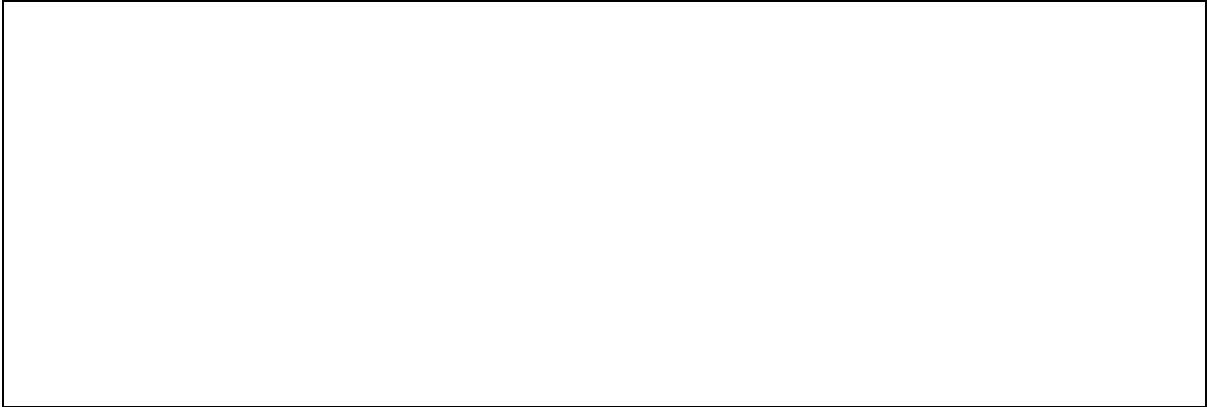|	
4. What do you think the functions of national moderation (as conducted by NZQA) are?	

Prompt questions if needed:
 	
  a. Do you think there are any functions to do with quality assurance?
  b. If so, what are they?	

c. Do you think there are any functions that are to do with maintaining public confidence in the education and qualifications students receive? d. If so, what are they?	
e. Do you think there are any functions that are to do with providing professional learning opportunities to teachers/tutors and assessors? f. If so, what are they?	
5. Are the above functions fulfilled by NZQA's moderation?	

6. In your view, are there any **other** functions of national moderation?
  g. If so, what are those other functions?
  h. Are they fulfilled by NZQA's moderation?	

Questions about interviewee's background and the education organisation

7. What role do you have with respect to moderation of assessment?	
8. How long have you been at this organisation? In this role?	
9. How long have you been a teacher/tutor, assessor, or administrator in formal education? a. Have all of these roles involved moderation? If so, which ones?	
10. Which education organisations and education (or training) sectors have you worked in? a. Have you observed (or experienced) any differences in the functions of internal moderation in the different sectors? b. If so, what are they?	
11. Is your organisation moderated by NZQA? Has it been moderated by NZQA in the past?	
12. Is your organisation moderated by any other standard setting bodies/Quality Assurance Bodies (e.g. ITOs)? If so, how many? a. How closely do the functions of their moderation align with the functions of NZQA's moderation? In what ways?	
13. (If interviewee is from a **school**) a. Approximately how many students does your school have? b. In your experience, does the size of the school impact on the functions of moderation? If so, how? c. When were your school's last ERO and MNA visits? Do you think that either of these has this had any impact on your understanding of the functions of moderation? If so, what?	

14. (If interviewee is from an **ITP**) d. Approximately how many equivalent full-time students does your organisation have?	
e. Was your ITP registered as a secondary school (e.g. a tertiary high school) **before** the recent Youth Guarantee/Vocational Pathways policy initiative? f. If so, do you think this has had any impact on the functions of internal moderation? And what are they?	

g. When was your ITP's last EER visit (or ITPQ Academic Audit, if you haven't had an EER yet)?
    h. Do you think this has had any impact on your understanding of the functions of moderation?
    i. If so, what?	

15. (If interviewee is from a **PTE**)
    j. How would you best describe the nature of your organisation's main business? Prompt if needed, such as:
       i. short duration, limited credit-type courses—such as first aid or confined space training
      ii. foundation education (e.g. employment or life skills-type level 1-3 programmes)
     iii. longer duration 120 + credit programmes (e.g. focussed on a specific duration)	

k. Approximately how many equivalent full-time students do you have?	
l. Are your students fee-paying? m. Does your organisation receive any funding from the TEC?	
n. How many sites does your organisation have? (E.g. One site, several sites but only in one region, in several regions, or nationwide?)	

o. When was your PTE's last EER visit (or NZQA Audit, if you haven't had an EER yet)?
    p. Has this had any impact on your understanding of the functions of moderation?
    q. If so, what?	

Do you have any other comments you would like to make about the functions of moderation of assessment?

Thank you for your time, it really is appreciated.

4. Synopsis of stage one interview findings

Internal moderation (IM)	NZQA moderation (NM)

IM Quality control
 	
• Check:
   ◦ that assessment is fair
   ◦ that assessment materials assess approved outcomes
   ◦ that assessment materials assess approved content
   ◦ that assessor decisions are reliable, accurate and consistent
   ◦ that learners' assessment experiences in a programme are consistent
   ◦ that an organisation has robust assessment practices
   ◦ the quality of instruction
   ◦ the quality of graduates
   ◦ student performance
   ◦ the currency of taught curriculum
• Compliance with internal or external requirements
• Show where more quality assurance focus is needed
• Feeds into organisation's self-assessment
	NM Quality control
	
	• Continuum: 'Is effective' to 'Is ineffective'
	• Check that:
	   ◦ assessment materials assess the standard requirements
	   ◦ assessment decisions are accurate and consistent
	   ◦ assessments are fair
	• Comply with NZQA requirements
	• Signals where more quality assurance focus is needed
	• Provides evidence for:
	   ◦ EER result and Provider Category
	   ◦ Self-assessment

IM Maintain public confidence • Continuum: 'Is' to 'Is not' a function • Occurs via public report • Reputation of organisation, quality of training and graduates • Because required part of quality assurance as a registered provider	*NM Maintain public confidence* • Continuum of 'Is effective' to 'Is ineffective' • Occurs via public reports • Because part of quality assurance required from a registered provider • In learner achievement; organisation's reputation

IM Professional learning opportunities	NM Professional learning opportunities
- Channels via which opportunities are provided: - moderation feedback or result - professional conversations - one-on-one support - Focus of opportunities: - requirements of standard or course - assessment decisions - assessment materials - assessment practice - teaching and learning - context-specific application - Alert organisation to where professional development is needed	- Ineffective - Channels via which opportunities are provided: - moderation feedback or result - professional conversations - Focus of opportunities: - assessment material - interpretation of standard requirements - assessor decisions

IM Quality enhancement	NM Quality enhancement
- Inform reviews of: - Curriculum - Teaching - Assessment - Student performance	Informs practice

IM Other	NM Other
- Input into performance appraisals - Internal moderation occurring at different times has different functions	- Provides confidence to management / governance teams re assessment, teaching occurring in organisation - Informs performance appraisals - Strong sense that NZQA moderation takes a compliance approach

Emergent themes

- Terminology issues:
  - what TEO respondents call 'national moderation', school respondents call 'external moderation'
  - moderation that is conducted by another education organisation is called 'external moderation' by TEO respondents, but 'internal moderation' by school respondents
- The functions that NZQA moderation **could** have are different from the functions that NZQA moderation **does** have.
- Moderation provides a forum for collegial conversations about teaching, learning, and assessment, and can assist developing a community of shared understanding
- Continuum: in the breadth of focus of moderation
  - narrow focus—only on assessment
  - main focus on assessment, with broader bi-products (e.g., teaching and learning)
  - broad focus—encompasses teaching, learning, curriculum, assessment
- Different contexts in which moderation functions: assessment; learning programmes; staff, organisational processes & quality assurance; organisational reputation; public confidence
- Can moderation handle integrated, student-centric assessment approaches?
- Assessment standards that are not fit-for-purpose are problematic for assessment and moderation

5. Survey information sheet

Academic leaders' perceptions of the functions of moderation: A mixed methods study

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI
VICTORIA
UNIVERSITY OF WELLINGTON

Survey Respondent Information Sheet

Researcher: Anna Williams, School of Education Policy and Implementation, Victoria University of Wellington

I am a Masters student in the Masters of Education at Victoria University of Wellington. As part of this degree I am undertaking a research project leading to a thesis. The purpose of the project I am undertaking is to examine the perceptions of academic leaders in education organisations about the functions of internal and national/external moderation.

I am inviting academic leaders (managers) in Institutes of Technology and Polytechnics, Private Training Enterprises, and state schools (integrated and not integrated, that include senior secondary levels) who have oversight of, and responsibility for, internal moderation and national/external moderation to participate in this study. These academic leaders will be asked to complete an online survey, which should take no longer than 30 minutes to complete. The survey will be anonymous; however respondents will have the option of entering their name, education organisation and contact details if they are happy to be contacted regarding participation in a follow-up interview.

Should any survey respondent who has included their name feel the need to withdraw their data from the project, they may do so without question anytime before **31 August 2014**. However, those who have not included their name or contact details will be unable to withdraw, as their survey responses will be unidentifiable.

The survey responses will form the basis of my research project and analysed in aggregated form, and discussed in a written report on an anonymous basis. It will not be possible for you or your organisation to be identified personally. All material collected will be kept confidential. No person besides me and my supervisors, Dr Michael Johnston and Dr Robin Averill, will see the survey responses. The thesis will be submitted for marking to the School of Education and deposited in the University Library. It is intended that one or more articles will be submitted for publication in scholarly journals, and one or more presentations may be made at academic or professional conferences. Survey responses will be destroyed five years after the end of the project.

I am undertaking this research in the capacity of a private citizen; however, I am employed by the New Zealand Qualifications Authority in the Assessment and Moderation Services business unit, where I oversee the moderation of several systems of assessment standards. I am conducting this study completely outside of my role with NZQA.

The research has been approved by the Faculty of Education Human Ethics Sub-committee under delegated authority from the Victoria University Human Ethics Committee. If you have any ethical questions about this research, please contact Dr Allison Kirkman, Chair of the Human Ethics Committee, Victoria University of Wellington (Allison.Kirkman@vuw.ac.nz, phone 04 463 5676).

If you have any further questions or would like to receive further information about the project, please contact me at 027 449 0703 or Anna.Williams2@vuw.ac.nz, or either of my supervisors Dr Michael Johnston or Dr Robin Averill, at the School of Education Policy and Implementation at Victoria University – Michael at 04 463 9675 or Michael.Johnston@vuw.ac.nz, or Robin at 04 463 9714 or Robin.Averill@vuw.ac.nz.

Anna Williams

6. Survey glossary

Academic leaders' perceptions of the functions of moderation: A mixed methods study

Survey Terminology

Terminology used in this survey:

- **Education** refers to education and training
- **NZQA moderation** refers to the national/external moderation conducted by NZQA
- **Organisation** refers to Institutes of Technology or Polytechnics, Private Training Enterprises, and Schools
- **Students** refers to students, learners, candidates and trainees
- **Teachers** refers to teachers, tutors, trainers, instructors, and assessors
- **Teaching** refers to teaching, instruction, and delivery

7. Survey instrument

Academic leaders' perceptions of the functions of moderation

Q1 **Academic leaders' perceptions of the functions of moderation**

This research aims to examine what academic leaders with oversight of internal and national/external moderation believe the functions of moderation to be.

Please complete this survey if:

- you are a senior or middle manager in an Institute of Technology or Polytechnic, a Private Training Enterprise, or a state (integrated or not integrated) school that includes senior secondary levels, and:
- you have oversight of, and responsibility for, internal moderation and national/external moderation (conducted by the New Zealand Qualifications Authority) in your organisation.

This research is being conducted completely independently of NZQA, although the researcher is employed there.

Your responses will be anonymous or confidential.

To view the Survey Information Sheet, please click here: Survey information sheet
To view the terminology used in this survey, please click here: Terminology used in survey

The survey is made up of four sections:

1. Preliminary questions
2. Questions about internal moderation—that is, moderation conducted within education organisations
3. Questions about national/external moderation of achievement standards and/or unit standards, as conducted by NZQA
4. Questions about you and your education organisation

It should take you no longer than 30 minutes to complete.

Thank you for taking the time to complete this survey. It really is appreciated.

Q2 **Section 1: Preliminary questions**

Q3 What type of organisation are you currently an academic leader or manager in?
- ❍ Institute of Technology or Polytechnic (ITP) (2)
- ❍ Private Training Enterprise (PTE) (3)
- ❍ State school (that includes senior secondary levels; integrated or not integrated) (4)

Q4 Please list the main functions of moderation

[open field]

Q5 **Section 2: Internal moderation (moderation conducted within education organisations)**

Q6 Please rate each of the following in terms of its importance as a function of internal moderation:

	No importance (1)	Low importance (2)	Medium importance (3)	High importance (4)
To check that assessments are fair to all students (1)	❍	❍	❍	❍
To check the assessment materials are fit for purpose in terms of the approved specifications or Standard (2)	❍	❍	❍	❍
To check that the assessment materials assess the content they are meant to (e.g. approved or industry-current content) (3)	❍	❍	❍	❍
To check that assessor judgements about assessment evidence are accurate (4)	❍	❍	❍	❍
To check that assessor judgements about assessment evidence are consistent (5)	❍	❍	❍	❍

Q7 Please rate each of the following in terms of its importance as a function of internal moderation:

	No importance (1)	Low importance (2)	Medium importance (3)	High importance (4)
To help ensure that an organisation meets NZQA's rules and requirements (1)	○	○	○	○
To meet the internal requirements of an organisation (e.g. as set out in the organisation's policies or quality management system) (2)	○	○	○	○
To provide an evidence source to be considered as part of an organisation's Self-assessment (ITPs and PTEs), or Self-review (Schools) (4)	○	○	○	○
To give confidence to the management team and/or governing body in the quality of teaching and assessment in their organisation (5)	○	○	○	○
To provide information for performance appraisals (6)	○	○	○	○

Q8 Please rate each of the following in terms of its importance as a function of internal moderation:

	No importance (1)	Low importance (2)	Medium importance (3)	High importance (4)
To check that the assessment practices in an organisation are robust (1)	○	○	○	○
To check the consistency of students' assessment experience throughout a programme (2)	○	○	○	○
To check that the approved and taught curriculum is current (4)	○	○	○	○
To show where more quality assurance activity should be focused within an organisation (3)	○	○	○	○

Q9 Please rate each of the following in terms of its importance as a function of internal moderation:

	No importance (1)	Low importance (2)	Medium importance (3)	High importance (4)
To check that the teaching that students have received has adequately prepared them for assessment (1)	○	○	○	○
To check that students are achieving at the levels that the organisation deems they should be achieving at (2)	○	○	○	○
To check that students are well prepared for going into further education or employment (3)	○	○	○	○

Q10 Please describe any other quality assurance functions that you believe are offered by internal moderation

[open field]

Q11 Please rate each of the following in terms of its importance as a function of internal moderation:

	No importance (1)	Low importance (2)	Medium importance (3)	High importance (4)
To help to ensure that other education organisations, parents, and employers can have confidence in an organisation's education programmes and qualifications (1)	○	○	○	○
To maintain public confidence in the internal assessment conducted in an organisation (2)	○	○	○	○
To maintain public confidence in national qualifications (3)	○	○	○	○

Q12 Please make any comments you might have about internal moderation having a function in maintaining public confidence in education and qualifications

[open field]

Q13 Please rate each of the following in terms of its importance as a function of internal moderation:

	No importance (1)	Low importance (2)	Medium importance (3)	High importance (4)
To give feedback about how well the assessment materials assess achievement against the Standard or course (1)	❍	❍	❍	❍
To give feedback about how to amend the assessment materials in order to better assess achievement against the Standard or course (2)	❍	❍	❍	❍
To give feedback about the quality of assessor judgements (3)	❍	❍	❍	❍
To give feedback about how to improve assessor judgements (4)	❍	❍	❍	❍

Q14 Please rate each of the following in terms of its importance as a function of internal moderation:

	No importance (1)	Low importance (2)	Medium importance (3)	High importance (4)
To give guidance about how to interpret the requirements of the Standard or course (1)	❍	❍	❍	❍
To assist in the development of a shared understanding among relevant people (2)	❍	❍	❍	❍
To build the confidence of assessment designers or assessors (3)	❍	❍	❍	❍
To alert the organisation to where professional development is required (4)	❍	❍	❍	❍

Q15 Please rate each of the following in terms of its importance as a function of internal moderation:

	No importance (1)	Low importance (2)	Medium importance (3)	High importance (4)
To provide opportunities for learning through moderation feedback (1)	○	○	○	○
To provide opportunities for professional/collegial conversations (2)	○	○	○	○
To provide opportunities for learning through one-on-one support (3)	○	○	○	○

Q16 Please rate each of the following in terms of its importance as a function of internal moderation:

	No importance (1)	Low importance (2)	Medium importance (3)	High importance (4)
To provide learning opportunities about assessment practices (1)	○	○	○	○
To provide learning opportunities about the context- or client-specific application of the requirements of the Standard or course (2)	○	○	○	○
To provide learning opportunities about teaching and learning (3)	○	○	○	○

Q17 Please describe any other professional learning opportunities offered by internal moderation
[open field]

Q18 Please rate each of the following in terms of its importance as a function of internal moderation:

	No importance (1)	Low importance (2)	Medium importance (3)	High importance (4)
To inform curriculum reviews (e.g. its currency, relevance, or tailoring for particular students) (1)	○	○	○	○
To inform teaching (2)	○	○	○	○
To inform assessment review (3)	○	○	○	○
To inform reviews of student performance (4)	○	○	○	○

Q19 Please make any other comments you might have about the functions of internal moderation
[open field]

Q20 **Section 3: NZQA moderation (of achievement standards and/or unit standards)**

Q21 Please rate the current emphasis of NZQA moderation on each of the following functions:

	No emphasis (1)	Low emphasis (2)	Medium emphasis (3)	High emphasis (4)
To check whether or not the internal assessments of an organisation meet the requirements of the Standards, as set nationally (1)	○	○	○	○
To check that assessment materials are fit for purpose in terms of the requirements of the Standard (2)	○	○	○	○
To check that assessor judgements are accurate against the requirements of the Standard (3)	○	○	○	○
To check that assessments are fair to all students (4)	○	○	○	○
To provide information about whether or not content that is assessed is appropriate (5)	○	○	○	○

Q22 Please rate the current emphasis of NZQA moderation on each of the following functions:

	No emphasis (1)	Low emphasis (2)	Medium emphasis (3)	High emphasis (4)
To provide evidence for consideration in an organisation's Self-assessment (ITPs and PTEs), or Self-review (Schools) (1)	○	○	○	○
To provide an evidence source that contributes to an organisation's External Evaluation and Review, and Provider Category (ITPs and PTEs), or Managing National Assessment review (Schools) (2)	○	○	○	○
To monitor an organisation's compliance with NZQA's rules and requirements (3)	○	○	○	○
To give confidence to the management team or governing body about an organisation's teaching and assessment (4)	○	○	○	○
To provide information for performance appraisals (5)	○	○	○	○

Q23 Please describe any other quality assurance functions offered by NZQA moderation [open field]

Q24 Please rate the current emphasis of NZQA moderation on each of the following functions:

	No emphasis (1)	Low emphasis (2)	Medium emphasis (3)	High emphasis (4)
To help to ensure that education and qualifications are seen as credible by parents, employers and other education organisations (1)	❍	❍	❍	❍
To maintain public confidence in the internal assessment conducted by organisations (2)	❍	❍	❍	❍
To maintain public confidence in national qualifications (3)	❍	❍	❍	❍

Q25 Please make any comments you might have about NZQA moderation having a function in maintaining public confidence in education and/or qualifications
[open field]

Q26 Please rate the current emphasis of NZQA moderation on each of the following functions:

	No emphasis (1)	Low emphasis (2)	Medium emphasis (3)	High emphasis (4)
To provide opportunities for learning from the moderation results in moderation reports (1)	○	○	○	○
To provide opportunities for learning from the feedback in moderation reports (2)	○	○	○	○
To give feedback about how well the assessment materials assess achievement against a Standard (3)	○	○	○	○
To give feedback about how to amend the assessment materials in order to better assess achievement against a Standard (4)	○	○	○	○
To give feedback about the quality of the assessor judgements (5)	○	○	○	○
To give feedback about how to improve assessor judgements (6)	○	○	○	○

Q27 Please rate the current emphasis of NZQA moderation on each of the following functions:

	No emphasis (1)	Low emphasis (2)	Medium emphasis (3)	High emphasis (4)
To give guidance regarding how to interpret the requirements of a Standard (1)	○	○	○	○
To stimulate professional and collegial conversation (2)	○	○	○	○
To assist in the development of a shared understanding among relevant people (3)	○	○	○	○
To build the confidence of assessment designers or assessors (4)	○	○	○	○

Q28 Please describe any other professional learning opportunities offered by NZQA moderation
[open field]

Q29 Please rate each of the following in terms of its importance as a function of NZQA moderation

	No importance (1)	Low importance (2)	Medium importance (3)	High importance (4)
To inform an organisation's practice (1)	○	○	○	○
To give feedback about whether an assessment supports teaching and learning (2)	○	○	○	○
To inform reviews of Standards (3)	○	○	○	○

Q30 Please make any other comments you might have about the functions of NZQA moderation

[open field]

Q31 **Section 4: Questions about you and your education organisation**

Q32 What is your job title?

[open field]

Q33 What is your role in your organisation with respect to moderation?

[open field]

Q34 For approximately how many years have you been involved in academic management of moderation in education?   Please respond in numeric form (e.g. 4), not text form (e.g. four). If you have been involved for less than 1 year, please answer "0".

Q35 For approximately how many years have you been a teacher, assessor or administrator in formal education?   Please respond in numeric form (e.g. 4), not text form (e.g. four). If your response is less than 1 year, please answer "0".

Q36 Which different types of education organisations and education sectors have you worked in or for? Select all that are applicable.

- ❑ Early Childhood sector (1)
- ❑ Primary school sector (2)
- ❑ Kura kaupapa (3)
- ❑ Private secondary school(s) (4)
- ❑ State secondary school(s) (5)
- ❑ Institutes of Technology or Polytechnic(s) (6)
- ❑ Private Training Enterprise(s) (7)
- ❑ Government Training Enterprise(s) (8)
- ❑ Wananga (9)
- ❑ Industry Training Organisation(s) (10)
- ❑ Work-place Training (11)
- ❑ Education Review Office (12)
- ❑ Ministry of Education (13)
- ❑ New Zealand Qualifications Authority (14)
- ❑ Tertiary Education Commission (15)
- ❑ Other (16) _____

Q37 Has assessment within your organisation been moderated by NZQA (for achievement standards and/or unit standards) within the last three years?

- ❍ Yes (1)
- ❍ No (2)

Q38 Have you had experience of national/external moderation by Standard Setting Bodies other than NZQA (e.g. Industry Training Organisations)?

- ❍ Yes (1)
- ❍ No (2)

Answer If Have you had experience of national/external moderation by Standard Setting Bodies other than NZQA (e.g. Industry Training Organisations)? Yes Is Selected

Q39 How many other Standard Setting Bodies (e.g. Industry Training Organisations) have you had experience of moderation by? (Please indicate the number before the recent ITO mergers.)

- ❍ 1–2 (1)
- ❍ 3–5 (2)
- ❍ More than 5 (3)

Q40 **For respondents in a State school:**

Q41 What decile is your school?

- ○ Decile 1 (1)
- ○ Decile 2 (2)
- ○ Decile 3 (3)
- ○ Decile 4 (4)
- ○ Decile 5 (5)
- ○ Decile 6 (6)
- ○ Decile 7 (7)
- ○ Decile 8 (8)
- ○ Decile 9 (9)
- ○ Decile 10 (10)

Q42 Is your school classified as main urban, minor urban, or rural?

- ○ Main Urban (1)
- ○ Minor Urban (2)
- ○ Rural (3)

Q43 Is your school a Composite or Secondary school?

- ○ Composite School (1)
- ○ Secondary School (2)

Q44 Approximately how many students in total does your school have?   Please respond in numeric form (e.g. 50), not text form (e.g. fifty).

Q45 Approximately how many senior secondary students (years 11 and above) does your school have? Please respond in numeric form (e.g. 50), not text form (e.g. fifty).

Q46 For respondents in an Institute of Technology or Polytechnic:

Q47 Is your ITP a member of the Metro-group?

❍ Yes (1)
❍ No (2)

Q48 Approximately how many Equivalent Full Time Students (EFTS) does your ITP have?

Please respond in numeric form (e.g. 50), not text form (e.g. fifty).

Q49 Was your ITP also registered as a secondary school before the recent Youth

Guarantee/Vocational Pathways policy initiative?

❍ Yes (1)
❍ No (2)

Answer If What type of organisation are you currently an academic leader or manager in? Private Training Enterprise (PTE) Is Selected

Q50 **For respondents in a Private Training Enterprise:**

Q51 Which of the following best describes the nature of your PTE's main business? (Select

all options that apply as your main business.)

❑ Foundation education (e.g. levels 1–3, employment skills or life skills-type programmes) (1)
❑ Short duration limited credit-type courses (such as first aid training, or confined space training) (2)
❑ Longer duration 120 credit (or more) programmes (e.g. focused on a specific vocation) (3)
❑ Recognition of prior learning (4)

Q52 Approximately how many Equivalent Full Time Students (EFTS) does your PTE have?

Please respond in numeric form (e.g. 50), not text form (e.g. fifty).

Q53 Are your students fee-paying?

❍ Yes, the majority (or all) are (1)
❍ Some are (2)
❍ No, none (or almost none) are (3)

Q54 Please indicate whether your organisation:

❍ is single site (1)
❍ more than one site within a region (2)
❍ has branches or affiliated sites in more than one region (3)
❍ Other (4) _____

Q55 Please specify how many regions your organisation has branches or affiliated sites in

[open field]

Q56 In which subject areas or moderation systems has NZQA moderated standards for your organisation within the last three years?

[open field]

Q57 If you are happy to be contacted after the survey regarding participation in a follow-up interview, please complete the following:

Your name: (1)

Your education organisation: (2)

Your email address (3)

Q58 Thank you very much for taking the time to complete this survey.

8. Invitation to participate in survey example: ITP email

Notes:
- This invitation was emailed to each ITP using mail merge from the researcher's Victoria University of Wellington email address.
- The Survey Information Sheet was attached to the email.
- An equivalent but tailored invitation was sent to each PTE and School in the target populations.

Email subject line:
Research into academic leaders' perceptions of the functions of moderation

Email content:

Tena koe

I am a student in the Masters of Education at Victoria University of Wellington. I am examining what academic leaders with oversight of internal and national moderation understand as the functions of moderation.

Your Institute of Technology or Polytechnic is invited to participate in this research.

In this research, the senior or middle manager who has oversight of, and responsibility for, internal and national moderation (conducted by the New Zealand Qualifications Authority) is invited to complete an online survey.

Please forward this email to the appropriate academic leader within your institution.

The survey should take no longer than 30 minutes to complete.

The survey is anonymous (unless you choose to leave your contact details, in which case it is completely confidential).

To take the survey, please go to http://vuw.qualtrics.com/SE/?SID=SV_9KPh6ZY910BfN7D

Please complete this survey by **Friday 27 June, 2014**.

The Survey Respondent Information Sheet is attached for your information.

Many thanks in advance

Anna Williams

9. Reminder invitation example: PTE email

Notes:
- This email was sent by mail merge to PTEs who may not have already responded.
- The Survey Information Sheet was attached to the email.
- An equivalent but tailored invitation was sent to ITPs and Schools.

Email subject line:
Re: Research into academic leaders' perceptions of the functions of moderation

Email content:

Tena koe

Several weeks ago, I sent you an invitation to participate in my research into the perceptions of academic leaders about the functions of moderation. If you or someone within your organisation has already completed the survey, thanks so much.

If you haven't, I would again like to invite your Private Training Enterprise to participate in this research.

As explained in my previous email, I am a student in the Masters of Education at Victoria University of Wellington.

In this research, the senior or middle manager who has oversight of, and responsibility for, internal and national moderation (conducted by the New Zealand Qualifications Authority) is invited to complete an online survey.

Please forward this email to the appropriate academic leader within your organisation.

To take the survey, please go to http://vuw.qualtrics.com/SE/?SID=SV_9KPh6ZY910BfN7D

The survey should take no longer than 30 minutes to complete.

The survey is anonymous (unless you choose to leave your contact details, in which case it is completely confidential).

Please complete this survey by **Friday 27 June, 2014**.

The Survey Respondent Information Sheet is attached for your information.

Many thanks again

Anna Williams

10. Second reminder invitation example: School email

Notes:
- This email was sent by mail merge to Schools who may not have already responded
- The Survey Information Sheet was attached to the email.
- An equivalent but tailored invitation was sent to ITPs and PTEs.

Email subject line:
Survey still open: Research into academic leaders' perceptions of the functions of moderation

Email content:

Tena koe ano

A number of weeks ago, I sent you an invitation to participate in my research into the perceptions of academic leaders about the functions of moderation. If you or someone within your school has already completed the survey, thanks so much.

I have postponed the closing date of the survey until Friday 18 July, so if you haven't, I would again like to invite your school to participate in this research.

As explained in my previous emails, I am a student in the Masters of Education at Victoria University of Wellington.

In this research, the senior or middle manager who has oversight of, and responsibility for, internal and external moderation (conducted by the New Zealand Qualifications Authority) is invited to complete an online survey.

Please forward this email to the appropriate academic leader within your school.

To take the survey, please go to http://vuw.qualtrics.com/SE/?SID=SV_9KPh6ZY910BfN7D

The survey should take no longer than 30 minutes to complete.

The survey is anonymous (unless you choose to leave your contact details, in which case it is completely confidential).

Please complete this survey by **Friday 18 July, 2014**.

The Survey Respondent Information Sheet is attached for your information.

Many thanks again

Anna Williams

11. Coding framework: Internal moderation

Global code	Parent code	Child code
IM: Assessment quality		
	General assessment quality	
		Endorse
		Dissent
	Assessment component	
		Assessment-in-general
		Assessment materials
		Assessor judgements & grades awarded
	Moderation action	
		Checking assessment material quality*
		Checking assessor judgement quality*
		Improving assessment material & assessor judgement quality*
		Undefined action
	Assessment property	
		Assess requirements & at correct level
		Consistent assessment
		Fair assessment
		Robust quality
		Valid assessment
IM: Educational quality		
	Educational quality	
		General educational quality
		Checking educational quality*
		Improving educational quality*
	Teaching, programme & curriculum	
	Student learning, progress & achievement	
IM: Maintaining public & stakeholder confidence*		
	Endorse role	

	Indirect role	
	Dissent or qualifier	
	NZQA mod role	
	Not publicly visible	
IM: Organisational & professional learning & development*		
	Professional learning & development opportunities	
	Communities of practice	
	Organisational development opportunities	
	Opportunities to identify professional development needs	
	Focus of opportunities	
		Assessment & standards
		Moderation
		Teaching & learning
	Qualified	
	Other factors preclude opportunities	
	Dissent	
IM: Organisational quality assurance*		
	Part of organisation's QMS	
	Informs other processes	
	Requirement of NZQA/QAB	
	Confirmation	
	Opportunity to address IM findings	

* Code founded on factor-based code(s)

12. Coding framework: NZQA moderation

Global code	Parent code	Child code
NM: Assessment quality		
	General assessment quality	
		Endorse
		Dissent
	Assessment component	
		Assessment-in-general
		Assessment materials
		Assessor judgements & grades awarded
	Moderation action	
		Checking internal assessment quality*
		Improving internal assessment quality*
		Action undefined
	Assessment property	
		Assess requirements & at correct level
		Consistent assessment
		Fair assessment
		Robust quality
		Valid assessment
NM: Maintaining public & stakeholder confidence*		
	Endorsement of role	
	Dissatisfaction or qualifier	
	Dissent	
	Not publicly visible	
NM: Organisational quality assurance*		
	Compliance with NZQA requirements	
	Concern re NZQA role	
	Organisation's own self-assessment	

NM: Professional learning & development opportunities		
	Learning opportunities provided by NZQA moderation*	
		Endorse
		Neutral
		Dissent
	Other NZQA support, guidance	
		NZQA-run workshops
		Other support & guidance provided by NZQA
NM: Educational quality		
	Endorse	
	Qualify	
	Dissent	
Dissatisfaction with NZQA		
	NZQA moderation	
	Standards	
	Support & resources	

* Code founded on factor-based code(s)

13. Coding framework: Moderation-in-general

Global code	Parent code	Child code
MIG: Assessment quality		
	Assessment component	
		Assessment-in-general
		Assessment materials
		Assessor judgements & grades awarded
	Action moderation takes	
		Assure
		Evaluate
		Feedback & improve
		Undefined
	Assessment property	
		Assess requirements & at correct level
		Accurate assessor judgements
		Consistent
		Fair
		Reliable
		Robust quality
		Valid
MIG: Quality assurance		
	Quality assurance	
	Organisational quality assurance	
	Externally-set requirements	
	QA of qualifications	
	System-level QA	
MIG: Educational quality		
	Teaching, programme & curriculum	
	Student learning, progress, & achievement	
MIG: Professional learning and communities of practice		
	Professional learning opportunities	

	Communities of practice	
		Professional conversations
		Shared understanding & practice
MIG: Maintaining stakeholder confidence		
MIG: Dissatisfaction with NZQA or political environment		

14. Code inclusions lists: Internal moderation extract

Code	Inclusions	Further inclusions
IM: Assessment quality (general)	Assessment where not specifically **checking quality** of assessment material or assessor judgements, or **improving quality** of assessment materials or assessor judgements. (Refer relevant verbs) Assessment practices, systems & methodologies; ensuring national standards are met Consistency & comparability, benchmarking (if not specifically about assessment materials or assessor judgements—e.g. if "against national standards")	If about quality of **assessment materials or assessor judgements**, verb not specified, or verbs include: Approve, Confirm, Ratify, Validate, Assure, Ensure, Maintain, Uphold, Adhere
IM: Checking assessment material quality*	Content of survey items: Q6.3 Internal moderation: To check that the assessment materials assess the content they are meant to (e.g. approved or industry-current content) Q6.2 Internal moderation: To check the assessment materials are fit for purpose in terms of the approved specifications or Standard	Responses explicitly or specifically address checking the quality of **assessment material/resources/activities/tasks** Verbs include: Evaluate, Check, Assess, Examine, Benchmark against, Verify, Compare, Critique, Measure, Review, Determine (whether/if), Ascertain (whether/if) Includes against national standards, other organisations etc—if explicitly about assessment materials, & using above verb
IM: Checking assessor judgement quality*	Content of survey items: Q6.5 Internal moderation: To check that assessor judgements about assessment evidence are consistent Q6.4 Internal moderation: To check that assessor judgements about assessment evidence are accurate Q6.1 Internal moderation: To check that assessments are fair to all students (NOTE THAT THIS APPEARS HERE)	Responses explicitly or specifically address checking the **quality of assessor decisions/judgements/ marking/grades** Verbs include: Evaluate, Check, Assess, Examine, Benchmark against, Verify, Compare, Critique, Measure, Review, Determine (whether/if), Ascertain (whether/if) Including checking consistency against (or comparison with) standard, national standard, other organisations (so long as explicitly about assessor judgements/grades/grading/marks) Checking Fairness of assessment

15. Coding guidelines: Internal moderation extract

Situation	Coding approach	To illustrate	Example from data
Lack of specificity in response so cannot ascertain which factor code(s) most relevant	Code to broad (general, undefined) code, if relevant	"assure" & "ensure"—not clear if "check" &/or "improve" "assessments"—not clear if assmt materials or assr judgements	1. "Consistency amongst organisations"—to <u>Assmt quality</u> b/c not specifed if assmt material or assr judgements;
Response doesn't fit clearly with any one code	Code to code with closest underpinning concept		59: "Int mod is also used as a source of info for ext mod"—coded to <u>Org QA</u> b/c about providing a source of info for other QA processes
How to ascertain concepts that align with each factor code, and distinguish between the codes	Review survey items associated with each factor		51: "moderation has limitations in measuring the levels of achievement when using unit standards alone…"—coded to <u>Checking educational quality</u> b/c item 9.2 has Int mod with the function "To check that students are achieving at the levels that the organisation deems they should be achieving at"
Response to specific question appears to be relevant to other code	Code to specific question area also	Q12: Please make any comments about int mod having a function in maintaining public confidence"; response: "Consistency amongst organisations" (1)—code also to <u>Maintain public confidence</u>	
Different parts of response directly relevant to different codes	Code each part to relevant code		51. "There are key factors identified by int mod—teaching by the tutor and learning by the learners"—coded to <u>Ed quality</u>; "There are key factors… and evidence collection related to the US by the learner"—coded to <u>Assmt quality</u>.

16. Iterations of development of coding framework

(Moderation-in-general)

B ★

Assmt grading
Assmt materials + grading (assy judgment)

- assmt materials allow students to access all grades
- approp for specific context (assmt material)
- fair, valid, consistent, robust
- transparency of grades awarded
- comparable, benchmarking
- standard - check vs integ

(3) Integrity of Qualifications, National Qual System
— comparability

Share evaluate
Provide fdbk
Compare

(5) Educational quality

- teaching
  - taught curric → comparability
  - teaching conditions regarding
  - meeting of NZ curric & meaning
  - student achievement

With academic standards
let curric Lo's delivery

tog/ meanings of award-? Pos of
e.g. train grad profile
ensure students taught content that accessed

(4) Professional learning opps
PD

- community of organisational practice
  - have understanding opp.
  - PD awareness

(4) ↑ confidence of teachers

(4) Identify PD needs
myDA(?)

Provide confidence to
Stakeholders
+ maintain

LNZQA

(1) Provide confidence to stakeholders
+ maintain

(1) Political purpose
→ NZQA, Govt.
→ are different to public conf ??

(3)(2) Meet regulatory requirements (eg, NZQA, MoE, TEC)
(2) - compliance
(2) meet int QA reqmts. Part of int QMS

(1) Organisation
- provide confidence to
  - ithrs, users
  - mgmt

(2) Organisational
→ grad, etc from edorg, ...

(3) Whole mgmt what to calculate
(3) Natnl agreement stds

credibility of grades
integrity of " "

(1) ed from edorg ...

Evaluate (self-eval or ext eval)
Provide fdbk about
Improve : guide

'Ensure', 'assure',
'confirm', 'make sure'

check, ratify, verify, confirm, evaluate,
provide fdbk = allows improvement,
guides
(6)
to guide/inform pract
: for improvement
assist improvement

(1) Forum for
fdbk about Assmt quality
& quality

fdfbh
to plan, assmt
develop
code conversn; ...
assess plant

idea from
NZ st makers

→ assure
→ assure, assure (make sure), maintain
proper fdbk for improvement guide
(to assist improvement)
→ allow improvement

go ahead ?
(ratify/...)

(6)

Whole Assessment management
(assmt/practical) management
(parents/examine, examine conducted system)
SUBJECT ★

c.

Assessment Quality

Assure, ensure, maintain, uphold Assessment quality

To improve assmt quality

To improve assmt materials

Provide fdbk about assmt Quality

Provide fdbk about Judgments & grades

Provide fdbk about assmt materials

Assure, ensure eh assessments materials

Assmt against check, Evaluate, verify, assess, examine Assessment Quality

Assessment check, evaluate, ea Assessment materials

Assessment critic evaluate, ea Assesses Judgements & grades

assure, ensure eh assesses Judgements, grades

consistency accuracy fairness comparability

Differentiated bhn Int med + nat mod

Ext/nat Mod; external view

② Organisational Quality assurance of organisation internal QA

self-assmt

③ System-level QA ??

Feedback mechanism forum for feedback

about assmt quality

about Ed quality

Integrity & comparability of quals - natil Quals system

③

④ Professional learning + communities of Practice

Response of topic

Educational quality

⑤

⑤a Assure, ensure ed quality

⑤b Check, evaluate ed quality

⑤c give fdbk → to improve

Stakeholder confidence

①

-incl political purposes

17. Contextual framework for moderation