The development of summative assessment literacy:

An exploration of the experiences of beginner secondary science teachers in New Zealand

A thesis

submitted in fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Education

at

The University of Waikato

by

FRANCES EDWARDS

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2017

Abstract

Summative assessment has been identified as an essential element of effective pedagogy. In the New Zealand context, secondary teachers are responsible for summative assessment within their own classrooms and for summative assessment that contributes to standards-based high stakes school leaving qualifications for students. For this they need high levels of summative assessment literacy. However, there is a lack of evidence on the nature of summative assessment literacy and its development, especially in the New Zealand context.

A qualitative approach underpinned by an interpretivist research paradigm was used to investigate summative assessment literacy development in eight beginner secondary science teachers. The study tracked the teachers from their entrance to a graduate one-year initial teacher education programme until they had completed six months teaching in their first schools. Each teacher was interviewed at five points through this approximately 18 month period. At each interview they were asked to present artefacts that they felt provided evidence of their understanding and use of summative assessment. These interviews helped establish the beginner teachers' understandings of summative assessment and allowed them to reflect on what and how they were learning. A questionnaire was used three times during the study (at the beginning and at the completion of the teacher education programme and after six months of teaching). The researcher also observed the students within their university classes. The data was analysed initially using a thematic analysis. Subsequently, a rubric (the SALRubric) was developed to allow further analysis of the development of specific dimensions of summative assessment literacy over time.

In the study, summative assessment literacy was found to be an amalgam of knowledges. Teacher commentary indicated that personal and contextual factors acted as amplifiers and/or filters for what teachers chose to implement in practice. Personal factors included teacher prior knowledge, experience, and conceptions about assessment; beliefs about ethical matters; and emotional responses to assessment consequences. Contextual factors included beginner teachers'

interactions and experiences with influential others as well as wider influences such as school policy and practice, and national policy. Beginner teachers attributed their summative assessment development to particular people, learning activities and teaching experiences. The profile of development of their summative assessment literacy was largely idiosyncratic with all beginner teachers demonstrating shifts over time. Development was more evident following practicum experiences.

This thesis provides a comprehensive and nuanced view of summative assessment literacy and its development. It highlights the need for policy makers as well as those working with beginner teachers in initial teacher education and schools to take a broad perspective on the nature of summative assessment literacy. A shared understanding regarding its development would allow beginner teachers and those working to support them to better understand the complexity of the process and lead to programmes that better prepare teachers for the challenge of this most complex and important task. The SALRubric developed from this study is a useful tool for tracking summative assessment literacy that could be used by and for beginner teachers to track their development and help focus learning.

Acknowledgements

There are many people who have helped me during my study and I would like to acknowledge and thank them.

Firstly thank you to my family. My thanks and gratitude go to my parents, William and Gene Eccles, for your encouragement, support and true belief in your daughters. Next, to my own dear daughters, Sarah, Katie and Stephanie, thank you for your support in small and big ways through this long process. And to Richard, thank you so much for your quiet confidence in me, your patience and support which made this work possible. I have appreciated your willingness to share in every part of my PhD adventure.

To my supervisors, who have encouraged and challenged me constantly, thank you. Beverley, thank you for starting me on this PhD and for your wise counsel along the path until your retirement. Brownen, thank you for the inspiration and for sharing your expertise so freely. Elaine, thank you for your wise guidance and encouraging words. I have learnt so much from you all.

To the wider academic family who have inspired me and kept me going, thank you. Dawn and Lise, I want to acknowledge your support and enthusiasm for helping novice researchers. Your passion for research has been inspiring. To my FEDU colleagues also on the PhD journey, I have really appreciated the camaraderie as we experienced the process together. Corridor meetings and encouraging words have kept us all going.

To my writing group buddies: Jenny, Kerry, Jeanette and Margaret, thank you so much for the support and encouragement. For all the laughs through the ups and downs, and for keeping things real, thank you. It will be great to celebrate when we have all finished our PhDs.

And finally thank you to my participants for being so open and generous with your time. It was a privilege to be able to hear your stories and to see your development over the 18 months of this study.

Table of contents

List of Tables

List of Figures

Chapter 1: Introduction

This chapter introduces the study and my research interest related to the development of summative assessment literacy in beginning secondary science teachers in New Zealand. It describes the study rationale and significance, outlines the research aims and gives an overview of the rest of the thesis.

1.1   The researcher's interest in the study

My interest in summative assessment was piqued early in my teaching career. In my third year as a secondary science and mathematics teacher I was challenged by a parent about the lack of formal assessment in her daughter's class. As a young, probably naïve, and ambitious teacher I had taken assessment decision making into my own hands and had designed fewer, but in my mind better, summative tasks for her class, despite the expectation that I would follow the school's regular testing regime. However, I was confident I had a good understanding of her daughter's progress, and I was now challenged with justifying this to the parent. Looking back at this, my decision to break with tradition was, if not a little rebellious, based on my strong beliefs about assessment.

I started my career teaching in an education system where teachers were responsible for secondary school summative assessment in the classroom, but norm-referenced external examinations were used for all senior secondary qualifications. Consequently, much of the classroom assessment mirrored the examination style used for the qualifications, particularly at senior levels. In 2002 this qualifications system was replaced by a standards-based assessment system — National Certificate of Educational Achievement (NCEA) — which in turn affected assessment practice in secondary school classrooms. This standards-based system is still in place in New Zealand secondary schools. During the early years of the implementation of this new assessment system I worked as a secondary school advisor and professional development provider, responsible for providing support for teachers in a large geographical region. Much of my work related to helping teachers develop a better understanding of standards-based assessment. This led to my employment as a national moderator for the New Zealand Qualifications Authority (NZQA), a role which involved monitoring and

check-moderating the work of a number of moderators employed to check assessment quality in school summative assessments that related to NCEA qualifications. Through my involvement in this I was exposed to the challenges of standards-based summative assessment that teachers faced. I was also fascinated by the wide range of summative assessment tasks that were designed by teachers to assess students against particular criteria. My subsequent move back to school to work as a secondary school leader meant a change of focus, as I had responsibility for school assessment, which included NCEA. This led to my development of a better understanding of the implications of assessment policy in a school setting. Most recently my work as a tertiary initial teacher educator means I am again able to focus centrally on teacher development. Overall, as I reflect on summative assessment, looking at the issues from a range of perspectives, a number of questions are raised for me, particularly about how best to help teachers develop both the confidence and the capability to assess students well.

My involvement in a small research project which focused on the development of assessment capability in preservice secondary teachers through their initial teacher education (ITE) programme (Cooper & Edwards, 2013; Edwards & Cooper, 2012) confirmed the complexities for secondary teachers developing and applying assessment knowledge and skills. This indicated to me that a qualitative study that provided the opportunity for an in-depth investigation into science teachers' development of summative assessment literacy would be useful and timely. I wanted to be involved in a study that had the potential to provide constructive insights for teacher educators and teachers who support beginner teachers. I therefore embarked on this study to investigate the characteristics of summative assessment literacy, and track the development of summative assessment literacy in beginner teachers from when they commenced their initial teacher education (ITE) until they had completed the first six months of employment in their first teaching positions. The term *beginner teachers* is used in this thesis to denote teachers involved in an initial teacher education programme and their first year of employment.

1.2 Assessment and assessment literacy

Teaching is a multifaceted activity that is cognitively demanding, requiring the use of a substantial body of knowledge and skills in the dynamic environment of the classroom. During their work day teachers are constantly faced with making pedagogical decisions, some of them planned and considered, and some of them more spontaneous and in-the-moment (Clark & Petersen, 1986; Darling-Hammond, 2006a; Gess-Newsome, 2015). Assessment has been identified as an essential element of effective pedagogy and has its own body of expertise. Excellence in teaching and learning is inextricably linked to assessment (Hargreaves, 1997; Ministry of Education, 2008) as teachers' use of effective assessment practices has been linked to benefits in student learning, increased levels of student achievement and improvements in pedagogy (Abell, 2007; Black & Wiliam, 1998a; DeLuca & Klinger, 2010; Mertler, 2005). Effective assessment is dependent on the assessors' knowledge and ability to apply this in the range of settings that present themselves in school classrooms as part of their teaching practice (Bell & Cowie, 2001; Black & Wiliam, 2006; Gardner, Harlen, Hayward, Stobart, & Montgomery, 2010; James & Pedder, 2006; Smith, 2010; Steiner Engelsen & Smith, 2014). Over the last 30 years or so there has been an increasing recognition of the need for teachers to have a breadth of knowledge to inform their assessment decisions. Additionally, there has been recognition of the influence of the lens a teacher brings, including teacher conceptions, context and affect, which may also act as amplifiers or filters (Gess-Newsome, 2015) when classroom decisions are made. Given that assessment is an essential professional capability, it is therefore imperative that teachers develop what has been described as assessment literacy.

Stiggins (1991) initiated discussions around the need for teachers to have assessment literacy, in terms of an 'assessment literate' person. He described those who were assessment literate as having "a basic understanding of high- and low-quality assessment and are able to apply that knowledge to various measures of student achievement" (p. 535), or more simply "knowing the difference between sound and unsound assessments" (Stiggins, 1995, p. 241). More recently Stiggins (2014) has re-emphasised what teachers must do, that is "to develop truly effective schools, educators must understand how to gather dependable evidence

of student achievement and use the assessment process and its results either to support or to certify student achievement depending on the context" (p. 67).

When a sociocultural lens is applied to assessment literacy or assessment literacies, assessment is seen as a cultural activity incorporating dynamic social practices which occur within specific contexts (Adie, 2011; Gee, 2008; Gipps, 1999; Pryor & Crossouard, 2008; Willis, Adie, & Klenowski, 2013). This view shifts the focus of assessment literacy from a singular or fixed set of capabilities, focused on individuals with a single discourse, to that of multiple assessment discourses in operation in situated complex contexts (Hay & Penney, 2012; Willis et al., 2013). The following definition has been offered by Willis et al. (2013):

> Assessment literacy is a dynamic context dependent social practice that involves teachers articulating and negotiating classroom and cultural knowledges with one another and with learners, in the initiation, development and practice of assessment to achieve the learning goals of students. (p. 242)

For the purposes of this study, the definitions from Stiggins (1991, 1995) and from Willis et al. (2013) offer useful perspectives from which to consider the development of summative assessment literacy in secondary science teachers. Summative assessment literacy is part of teacher learning, which has been characterised as having both individual and sociocultural features. The use of two lenses allows for a fuller representation of teacher learning and development to be explored, in which individual knowledge and skills (using a cognitive lens) are seen to develop as they become members of a practice community (using a sociocultural lens) (Clarke & Hollingsworth, 2002). As Cobb (1994) explained, "learning should be viewed as both a process of active individual construction and a process of enculturation into the … practices of wider society" (p. 13). Other researchers in mathematics education and science education have also found benefits from the use of more than one lens when investigating teacher practice (Buxton, Salinas, Mahotiere, Lee, & Secada, 2015; Cobb, 2007; Driver, Asoko, Leach, Mortimer, & Scott, 1994; Hodge & Cobb, 2016; Sfard, 1998; Varela, Thompson, & Rosch, 1991).

The literature concurs that assessment literate teachers are better able to make fully informed decisions about assessment, directed to a large extent by their

pedagogical knowledge about curriculum, learning and assessment, rather than by following a list of technical prescriptions. This implies decisions made in one context, for example, for an individual student, may be different from decisions made in a different context and for a different student (Steiner Engelsen & Smith, 2014).

1.3   Teacher development of assessment literacy

Assessment has been described as "the most challenging, theoretical, skill and knowledge area in education" (Brown, Irving, & Keegan, 2008, p. 12). Accordingly, assessment literacy has been identified as a core professional requirement for teachers in a wide number of educational settings (DeLuca, LaPointe-McEwan, & Luhanga, 2016a). Teacher education programmes for both pre-service and in-service teachers often include elements which aim to help teachers learn to assess well, as they develop their capacity to work more effectively within their educational contexts. Much current scholarship points to the effects of ITE programmes in developing teacher confidence and assessment literacy (DeLuca & Klinger, 2010; Edwards & Cooper, 2012; Eyers, 2014; McGee & Colby, 2014). However, some studies indicate inconsistent or weak assessment literacy development in ITE programmes (Deneen & Brown, 2016; Maclellan, 2004; Volante & Fazio, 2007). In particular, teachers' pre-existing beliefs and conceptions of assessment have been found to be robust and not easily changed (Deneen & Brown, 2016; Ecclestone & Pryor, 2003). Many novice teachers report being inadequately prepared and feeling challenged by aspects of summative assessment such as assessment task design, the need to ensure equity for all learners, how to mark fairly and how to communicate assessment results (Koh, 2014; Lyon, 2013a; Smith, Corkery, Butler, & Calvert, 2013). Unfortunately, the constraints within ITE programmes mean that only a rudimentary level of assessment literacy can be expected of beginner teachers (DeLuca, Klinger, Searle, & Shulha, 2010; Deneen & Brown, 2016). In light of this, scholars argue for a closer scrutiny of what and how ITE can best be delivered to enhance inservice teacher education and support ongoing teacher assessment literacy development (Black, Harrison, Hodgen, Marshall, & Serret, 2010; Hill, Cowie, Gilmore, & Smith, 2010; Struyven, Dochy, & Janssens, 2008).

There are, however, very few investigations into the development of summative assessment literacy in the New Zealand context. Recent studies confirm that teachers increase their understanding of assessment over time and leave ITE programmes at emergent levels, still requiring considerable support when they commence their first teaching position (Edwards & Cooper, 2012; Eyers, 2014; Smith, Corkery, Buckley, & Calvert, 2013; Smith, Hill, Cowie, & Gilmore, 2014).

1.4 Assessment in the New Zealand context

New Zealand's education policy is distinctive in that it operates within one of the most devolved school systems in the world, where teachers in self-managing schools are able to work with considerable freedom and flexibility to develop teaching programmes that best suit their students within the broad framework of the *New Zealand Curriculum* (Nusche, Laveault, MacBeath, & Santiago, 2012). Because of the nature of New Zealand's self-managed schools there are no nationally mandated tests or examinations in secondary education so schools can tailor an assessment regime that suits their students (Brown, Irving, & Keegan, 2014). New Zealand national assessment policy has placed substantial significance on teachers' use of formative assessment practices over a number of years (Crooks, 2002, 2011; Nusche, Laveault, MacBeath, & Santiago, 2012). All teachers are expected to view all assessment as a means to improve students' learning and teachers' teaching.

In senior secondary schools though, there is less focus on formative assessment. Summative assessment is mostly shaped by the needs for NCEA both in senior secondary, Years 11–13, where students work towards the standards-based NCEA qualifications, and in junior secondary, Years 9–10, where students are being prepared for future NCEA-style assessment. NCEA is gained through standards-based summative assessment and is the main secondary school leavers' qualification. NCEA requires secondary teachers to have a high level of assessment literacy, as it is achieved through the accumulation of credits gained by success in standards that are externally examined and marked, and standards that are internally assessed by classroom teachers. Individual schools and classroom teachers decide on the range and proportion of internally-assessed versus externally-assessed standards offered to their students. The national pattern

is that a higher proportion of NCEA credits are gained through internally-assessed standards than those that are externally assessed (New Zealand Qualifications Authority (NZQA), 2017a). Consequently, almost all senior secondary school teachers are involved in designing, administering and marking a wide range of internal summative assessments.

New Zealand teachers make decisions about student readiness, choose what standards are assessed, design or modify the assessment tasks, decide when the assessments will take place, and mark and grade assessment tasks. Quality assurance for this process is ensured through a sample of all teacher assessments being externally moderated by NZQA. NZQA provides feedback to teachers on summative assessment tasks and graded work that has been selected for moderation, as well as other standard-specific tasks that teachers voluntarily submit for critique. Therefore, in New Zealand, teachers play a central role in decision making with respect to summative assessment processes in ways not commonly seen elsewhere in the world (Crooks, 2011). New Zealand's assessment approach has been characterised by a high level of trust in schools and teachers' professionalism (Nusche et al., 2012). This educational context means that summative assessment literacy is particularly important for New Zealand secondary teachers.

1.5   Significance of research topic

Summative assessment literacy is a fundamental requirement for all teachers working in New Zealand schools. Its development through ITE programmes and in the early stages of teachers' careers is particularly vital. Teacher summative assessment literacy is not merely attained through teachers' attention to a basic list of technical skills, but is instead complex. Many factors influence teacher assessment decision making and application of what they know about assessment in their classrooms (Shepard, Hammerness, Darling-Hammond et al., 2012). Research investigating the nature of assessment literacy and how it is developed over time is scarce, and there have been calls for further research into this area internationally (DeLuca, Chavez, Bellara, & Cao, 2013; DeLuca, LaPointe-McEwan, & Luhanga, 2016b; Hill, Smith, Cowie, Gilmore, & Gunn, 2013; Lyon, 2012; Mertler, 2009; Moss, 2013).

In New Zealand secondary schools all teachers have a responsibility to summatively assess student learning as an integral part of their classroom work. In secondary schools in particular, many of the decisions and practices pertaining to summative assessment for qualifications and reporting are made by classroom teachers. There is little empirical research investigating the characteristics of summative assessment literacy of beginner teachers and the development of summative assessment literacy for New Zealand secondary school teachers. Because of New Zealand teachers' hands-on involvement in high stakes assessment, it is critical to gain a clear understanding of the characteristics of summative assessment literacy, including the contributors to its development, as teachers start their careers. The results of this study will be highly relevant to those involved in teacher education and beginner teacher mentoring in high stakes testing contexts in New Zealand as well as to teachers embarking on the profession.

The aim of this study was to investigate New Zealand secondary science beginner teachers' summative assessment literacy development from the start of their ITE programme until six months after they had commenced working in their first teaching positions. In particular, the characteristics of their summative assessment literacy, and how it develops over time, was the focus of this study. Of interest also was the identification of what beginner teachers felt contributed to the development of their summative assessment literacy. Given that there are very few longitudinal studies of assessment literacy development, and none focusing on summative assessment in the context of New Zealand secondary teachers, this research is timely to address this gap.

1.6   Research design

This was a qualitative study in which eight beginner science teachers were involved as participants through their ITE programme and into the first year of their employment. Data was collected via the use of a questionnaire and through a number of individual interviews conducted over time. Observations of aspects of their ITE learning were made, and at each interview the beginner teachers were asked to identify and supply artefacts that represented aspects of their summative assessment.

The conceptual framework guiding this study was based on two models: the Model for Science Teacher Assessment Literacy (Abell & Siegel, 2011) and the Model of Teacher Professional Knowledge and Skill including Pedagogical Content Knowledge (PCK) (Gess-Newsome, 2015). The first model was derived from empirical work on science assessment, and supported by the theoretical frameworks of Magnusson, Krajcik, and Borko (1999) and Pellegrino, Chudowsky, and Glaser (2001). The second model was derived from the synthesis of a large body of research on PCK in science education. PCK is defined as the knowledge that is developed by teachers to help others learn. This second model emphasises the inter-relationships between teacher content knowledge, professional knowledge and their classroom practice. It acknowledges the effects of beliefs, orientations, prior knowledge and context and posits that they act as amplifiers and filters for teacher decision making relating to classroom practice. Aspects of teacher summative assessment literacy were investigated, and a rubric was designed in order to track teacher summative assessment literacy development over time.

1.7   Structure of thesis

This thesis is organised into nine chapters. Chapter 1 provides an introduction to the research focused on New Zealand beginner secondary science teachers' summative assessment literacy development, the rationale for the research topic and its significance, and the conceptual framework adopted.

Chapters 2 and 3 review literature relevant to this study. Chapter 2 reviews literature that focuses on assessment and assessment literacy. Chapter 3 reviews literature that focuses on teachers' development, and in particular their development of assessment literacy. This chapter concludes with a summary and highlights the gap in the literature on developing beginner secondary science teachers' summative assessment literacy, leading to the research questions.

Chapter 4 outlines the research process and associated procedures. A description of the generation of a summative assessment literacy rubric (SALRubric) designed to track assessment development over time is provided, and its use in the analysis of all data gathered over the time period of the study is explained.

Evidence of trustworthiness in the study and considerations of ethical issues are also presented.

Chapters 5 to 7 report on the research findings. Chapter 5 reports findings concerned with teacher summative assessment knowledge, detailing the wider landscape of knowledge across beginner teachers. This is provided in five sections: knowledge of purposes of assessment, knowledge of what to assess, knowledge of assessment strategies and task design, knowledge of summative assessment interpretation, and knowledge of quality summative assessment.

Chapter 6 presents the findings relating to summative assessment literacy development over time through the use of the SALRubric. The developmental patterns of individual dimensions of summative assessment literacy are presented, followed by a description of the development of dimensions for individual teachers, including three documented examples. Next, the contributors to summative assessment literacy development identified by beginner teachers are presented. In particular, people, learning activities and teaching experiences were identified as contributing to beginner teachers' summative assessment literacy development.

Chapter 7 describes the amplifiers and filters that were found to influence beginner teachers' approach to knowledge use and their reported assessment decision making and actions. In addition the effects of sociocultural context, teachers' conceptions, ethical matters, and teacher emotional responses are described.

Chapter 8 discusses the findings from Chapters 5 to 7, under three headings which directly respond to the research questions. The first section discusses the characteristics of summative assessment literacy. The second section discusses the development of summative assessment literacy over time, and the third section discusses the contributors to the development of summative assessment literacy.

Chapter 9 summarises the key points of the study. The major implications for policy makers, ITE lecturers, schools and beginner teachers are discussed. An outline of the limitations of the study and suggestions for further research are also presented.

Chapter 2: Assessment

The purpose of this literature review is to provide a theoretical base and establish where this study sits within the body of knowledge relating to summative assessment and the development of teachers' summative assessment literacy. The literature review is organised into two chapters. This first chapter, Chapter 2, focuses on assessment, and in particular summative assessment, followed by Chapter 3 which focuses on teacher knowledge and the development of summative assessment literacy. Chapter 2 begins by framing assessment (Section 2.1) by considering international assessment trends, purposes of assessment with a focus on summative assessment quality and impact of summative assessment. Section 2.2 reviews literature on summative assessment in the New Zealand context including research into the assessment practices for NCEA. Next is a review of definitions of teachers' assessment literacy found in the literature including the definition of and rationale for assessment literacy in this study (Section 2.3).

2.1 Framing assessment

In the broadest sense, assessment, as a process of gathering evidence from which to make judgements and base decisions, is a part of the everyday world of all people. It occurs naturally as part of life. Everyday people make judgements and decisions based on the weighing up of information they see as relevant for the purposes of their decisions. The use of a range of forms of assessment within education is an extension of this process and is integral to teaching and learning. Without assessment, teachers and students would have no way of knowing whether or what learning is taking place (Broadfoot, 2012). However, assessment is "unquestionably one of the teacher's most complex and important tasks" (Moss, 2013, p. 235). It requires specialist knowledge and understanding. This section firstly reviews international assessment trends (Section 2.1.1). Then the purposes of assessment are reviewed (Section 2.1.2) with a particular focus on summative assessment (Section 2.1.3). This is followed by a section on quality in summative assessment design and use (Section 2.1.4) and finally a section on the impact of assessment (Section 2.1.5).

2.1.1   International assessment trends

International research and reviews of evaluation and assessment policy and practice across a number of countries reveal common patterns and trends in the use of assessment. Recent Organisation for Economic Co-operation and Development (OECD) reports highlight the increasing focus placed by governments and education policymakers on the formal assessment for a range of feedback and accountability purposes (OECD, 2013). Alongside this focus on externally mandated and managed summative assessment there is an increasing focus on involving classroom teachers in summative assessment of learning including in school leaver qualifications. There is also a growing awareness of the benefits of formative assessment internationally, with accompanying promotion of its use.

The increased prominence of formal school assessment in OECD countries is linked to increased demands for accountability in terms of school and teacher effectiveness, efficiency and quality (Nichols & Harris, 2016; Nusche et al., 2012; OECD, 2013), often interpreted as part of a neoliberal agenda (Connell, 2013; Davies & Bansel, 2007; Hill & Kumar, 2012; Hursh, 2007; Torrance, 2015; Watkins, 2007). International trends include an increase in governments monitoring progress within schools by way of school evaluation, teacher appraisal, and wider use of performance data (OECD, 2013) and a rising reliance on educational standards (DeLuca et al., 2016a; Flockton, 2012; Looney, 2011; Nusche et al., 2012; OECD, 2013). In many countries (for example, USA, England, New Zealand, Canada) it has been found that the increased political pressure demanding accountability measures is negatively affecting teaching and learning in classrooms through the way learning and success must be measured and reported (Gardner, 2006; Gipps, Brown, McCallum, & McAlister, 1995; Nichols & Harris, 2016; O'Neill, 2013; Thrupp & Easter, 2012).

Often the use of the results from student summative assessment is outside school and teacher control. Conflicting agendas exist when assessment designed for one purpose (e.g., learner/school improvement) is also used for other purposes (e.g., accountability). In several countries, for example the USA, France, United Kingdom, Canada, China, Japan and Korea, centrally designed standardised tests and examinations are the predominant tools used to gather information about what

students know and can do, but in some cases the data these generate has been used as an indicator of teacher quality or school effectiveness (Carless, 2007; Klenowski & Wyatt-Smith, 2014; Nichols & Harris, 2016; Stobart, 2008). Other examples include the use of national external standardised tests or national standards results for school ranking, and the release and publication of these school rankings. This is despite validity concerns around the use of such measures for these purposes (Black, 2013a; Stobart, 2008). The conflict inherent in teachers not being able to control or properly explain summative assessment data can cause tension when unintended consequences result.

Recently, an international trend has emerged to involve teachers in national assessment, giving them some control over and more responsibility in the assessment process. In many countries (e.g., Scotland, Wales, Northern Ireland, England, New Zealand, Australia) there is a move to depend more on assessment completed by classroom teachers (Crooks, 2011; Gardner, Harlen, Hayward, Stobart, & Montgomery, 2010; Panizzon & Pegg, 2008). This does not mean that standardised or externally set and marked tests have disappeared entirely from any of these countries' assessment landscapes at senior secondary school, particularly for exit qualifications, but it does mean that a wider range of assessment tasks are now being considered as valid and reliable options. This enables more agency on the part of the teacher and can lead to a more robust assessment of students' learning (Assessment Reform Group, 2006; Harlen, 2005a; Mansell, James, & Assessment Reform Group, 2009). It also allows for a range of innovative summative assessment strategies to be used; for example, the incorporation of digital technologies into assessment (Darling-Hammond, Herman, Pellegrino et al., 2013; Otrel-Cass, Cowie, & Glynn, 2009; Venville, Sheffield, Rennie, & Wallace, 2008). The professional judgement of teachers is relied upon in these contexts with systems set in place for quality assurance (Flockton, 2012; Wyatt-Smith, Klenowski, & Gunn, 2010). It is argued that the advantage of using teachers' judgements is that teachers are better placed to make informed judgements because of their greater knowledge of students, and so the validity and reliability of results is enhanced (Black, 2013a).

Because of the trend internationally for classroom teachers to have more responsibility for assessment, it is contended that the inclusion of both formative

and summative assessment literacy and competency development are necessary elements in teacher education (Black, Harrison, Hodgen, Marshall, & Serret, 2010; Boyd & Hayward, 2010; DeLuca & Johnson, 2017; Gardner, Harlen, Hayward, & Stobart, 2011). This argument has also been promoted in the New Zealand context (Hill, Cowie, Gilmore, & Smith, 2010; Jones & Buntting, 2013). Although the accountability pressures on New Zealand teachers are mild by international standards in that New Zealand has no mandated whole cohort testing regimes using standardised tests, and teachers' jobs are not dependent on their students' summative assessment data, this situation is being challenged. Crooks (2011) observed, when speaking of New Zealand, that this situation is "often challenged through political and media criticism … [which] leads to pressure to introduce more intrusive accountability measures. These could shift the balance of teachers' practice away from assessment for learning and towards assessment of learning" (p. 77). New Zealand secondary teachers have an important role in NCEA assessment for exit qualifications, so for these teachers summative assessment is significant. In the context of this study, it is important therefore that teachers gain an awareness of the external accountability pressures that exist for schools, and the ways in which summative assessment is used for accountability purposes. The development of summative assessment literacy enables secondary teachers to be better able to navigate their responsibilities to students, their families and communities, as well as to NZQA and external agencies.

Another important trend internationally is the increasing prioritisation in policy for the use of formative assessment, otherwise known as assessment for learning, by classroom teachers because of its capacity to support learners and raise student achievement (Black & Wiliam, 1998a; Cowie, 2013; National Research Council, 2010; OECD, 2013). A number of jurisdictions around the world now demand that classroom teachers assess students formatively and use the information they gather to better inform their ongoing planning and teaching. A recent review of international standards and measures for teacher assessment literacy standards and measures from five English speaking countries found that in early documents (1990–1999) summative assessment was the primary focus (DeLuca, LaPointe-McEwan, & Luhanga, 2016). A central foci on assessment purposes, assessment processes, communication of results and fairness was evident in standards for

assessment literacy from the 1990s until the present time. From 2000 onwards formative assessment became a focus and is now a dominant theme in a number of countries' assessment standards for teachers. Because assessment literacy is an established core professional requirement, this review recommended that future assessment literacy measures for teachers reflect contemporary understandings of assessment as reflected in teacher standards.

In summary, there are conflicting agendas at work in the educational assessment arena in the current international environment. On one hand, political pressures are putting an increasing accountability burden on schools and teachers to demonstrate impact, leading to calls for standardisation of assessment tools to allow for easy comparison between students, teachers and schools. Countering this is the growing body of research evidence pointing to the benefits of teachers being more centrally involved in assessment and the pivotal role of formative assessment. Internationally a number of countries are responding by developing standards for teachers that promote support and to account for teacher learning, and that involve the use of both formative and summative assessment by teachers. For teachers a professional knowledge of assessment and an understanding of its various purposes and practices is essential in this complex context.

2.1.2   Purposes of assessment

This section provides a review of the purposes and practice of assessment. Firstly, the terminology used to discuss purposes is set out (Section 2.1.2.1), followed by a description of formative assessment (Section 2.1.2.2) and an overview of aspects of summative assessment (Section 2.1.2.3).

2.1.2.1   Terminology used for assessment purposes

Assessment can be viewed as having two overall purposes: to inform decisions about learning progress and to report on what learning has been achieved (Harlen, 2007; Harlen & James, 1996; Sadler, 1989; Stiggins, 2007). That is, the information it generates has dual purposes. It can be used for improvement and for accountability, both of which are important in current educational settings (Brown et al., 2008; Gardner, Harlen, Hayward, Stobart, & Montgomery, 2010). A number of terms have been used to describe these assessment purposes as follows: assessment to inform decisions about learning progress (termed formative assessment, assessment for formative purposes, or assessment *for* learning), and

assessment to report on what learning has been achieved, usually occurring at the end of a block of learning (termed summative assessment, assessment for summative purposes or assessment *of* learning) (Black, 1998; Brookhart, 2011; Millar, 2013; Pellegrino et al., 2001; Scriven, 1967; Ussher & Earl, 2010).

Newton (2007) and Taras (2005) question the usefulness of terms like 'summative', 'formative', and 'purpose' to describe assessment. Both conclude that summative assessment has no purpose; it is simply an assessment judgement whereas formative assessment "can only meaningfully characterize a type of use to which assessment judgements are put" (Newton, 2007, p. 156). Other scholars concur that at times the terms 'formative' and 'summative' are less than helpful as people tend to attach these terms to the assessment tasks; for example, labelling a task as a 'formative task', when instead the terms aim to categorise assessment purpose. These terms "apply not to the assessment themselves, but to the functions they serve" (Wiliam & Black, 1996, p. 538). It is generally agreed that the key focus should be on the purpose of an assessment task, not on the assessment task itself (Newton, 2007; Ussher & Earl, 2010; Wiliam & Black, 1996). Teachers themselves are often in a position to decide of the functions of assessments they use, whether these be to inform pedagogical decisions or report on student achievement. When Crooks considered the functions of assessment in New Zealand schools he remarked "there is no clear demarcation line between assessment for development and assessment for accountability" (2002, p. 241), indicating that some assessment data can be used for a range of purposes.

Although there is debate on terminology and the ways that the purposes of assessment are categorised, most authors agree that there is overlap and interaction between assessment purposes, and that assessment purposes cannot be viewed in isolation from one another as all can contribute to effective pedagogy (Corrigan, Buntting, Jones, & Gunstone, 2013; Ministry of Education, 2011a; Shute & Kim, 2014). However, different assessment purposes require different assessment processes, so mixing purposes may not be helpful in ensuring that any of them are well serviced (Earl, 2013; Gipps, 1994). It is important that teachers are well placed to make decisions with respect to the assessment purposes they have and tasks they use. They need to be able to judge and make choices about

which assessment purpose and associated task is most appropriate in any given context (Pellegrino et al., 2001; Steiner Engelsen & Smith, 2014; Stiggins, 1995).

For the purposes of this thesis, the terms formative assessment and summative assessment are used to describe assessment purposes. Essential teacher professional knowledge is considered to include knowledge of both of these purposes. Each of these is described in more detail in the following sections.

2.1.2.2   *Formative assessment*
For the purposes of this study the term formative assessment is used synonymously with assessment for learning and assessment for formative purposes. Although formative assessment is not a focus for this study, it is briefly reviewed here because of its centrality in New Zealand education policy.

Black and Wiliam (2009) developed the following definition for formative assessment, placing it within a framework of broader theories of pedagogy:

> Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited. (Black & Wiliam, 2009, p. 9)

Formative assessment is seen as a major component of classroom teaching practice, essential to good teaching and inextricably linked to it (Bell, 2007; Black & Wiliam, 1998b; Carr et al., 2005; Crooks, 1988; National Curriculum Task Group on Assessment and Testing, 1988). As a result of the benefits attributed to the use of formative assessment, it has become a priority for teacher professional learning in a wide range of countries including New Zealand (Absolum, 2006; Absolum, Flockton, Hattie, Hipkins, & Reid, 2009; Brookhart, 2011; Crooks, 2011; Gardner et al., 2010; Taras, 2009). Professional development for formative assessment has been well funded in New Zealand for the last 20 years, resulting in it being part of the 'landscape' of teaching in New Zealand (Crooks, 2011).

Formative assessment has an improvement orientation as teachers assess the learning progress of their students and then adjust their instructional activities, and/or students adjust the ways they are trying to learn, in response to the progress they are making (Black & Wiliam, 1998a; Popham, 2000, 2009). For this reason

formative assessment is most likely to be ongoing, carried out during the learning and teaching process, informing pedagogical decisions. Key pedagogical practices linked to formative assessment that have been shown to assist with learning have been identified. For example, Wiliam and Thompson (2007) have drawn on earlier work by Black et al. (2003), together with the framework of Ramaprasad's (1983) three key processes in learning and teaching, to develop a framework of aspects of formative assessment. From this, Wiliam and Thompson (2007) suggested that formative assessment could be conceptualised as consisting of five key strategies

1. Clarifying and sharing learning intentions and criteria for success;
2. Engineering effective classroom discussions and other learning tasks that elicit evidence of students understanding;
3. Providing feedback that moves learners forward;
4. Activating students as instructional resources for one another; and
5. Activating students as owners of their own learning.

(Wiliam & Thompson, 2007, p. 64)

The overarching idea for these five strategies is that the evidence of student learning is used to adapt or adjust teaching and learning activities for the benefit of the students.

Assessment *as* learning, or "learning-oriented assessment" (Carless, 2007, p. 57), is sometimes seen as a subset of formative assessment. It places an emphasis on the role of the learner in assessment. Earl (2013) emphasises the importance of the learner's active involvement in the development of metacognitive skills. She proposes that by engaging deeply with the assessment process learners can "become comfortable with reflection and the critical analysis of their learning" (Earl, 2013, p. 28). Earl envisages classrooms where assessment is seen *as* learning as being effective in empowering students. Carless (2007) introduced the term "learning-oriented assessment" (p. 57) to help readers focus on the learning aspects of assessment. Both Earl (2013) and Carless (2007) emphasise the benefit in prioritising the development of the students as owners of their own learning.

The conceptualisation of students as agents for their own learning and the learning of others has led to a shift in policy direction for a number of countries. In New Zealand, both the New Zealand curriculum and Ministry of Education position papers on assessment emphasise the agency of students as well as teachers, and

call for assessment capability development in both groups (Ministry of Education, 2007, 2008, 2011a). The New Zealand curriculum states that

> the primary purpose of assessment is to improve students' learning and teachers' teaching as both student and teacher respond to the information that it provides. With this in mind, schools need to consider how they will gather, analyse, and use assessment information so that it is effective in meeting this purpose. (Ministry of Education, 2007, p. 40)

The New Zealand Ministry of Education position paper on assessment extends this to assert that all participants of the schooling system need to develop their assessment capability and that assessment capability is crucial to improvement. The challenge to develop the assessment capability in students has been discussed by Booth, Hill, and Dixon (2014). They describe conditions that need to be provided by teachers to nurture assessment capability development in students. The ability to provide such development presupposes assessment capable teachers (i.e., these teachers have learnt about assessment and are well able to create a context for and implement strategies that support student assessment capability and agency).

Formative assessment is not the focus of this thesis, but given its prioritisation in New Zealand's educational policy and the history of professional development available to teachers over many years, it is important to acknowledge that in the New Zealand context the idea of using assessment information formatively is widely accepted as the basis of teachers' practice. The centrality of the focus on the formative use of all assessment data has implications for teachers' classroom summative assessments in the sense that data collected for summative assessment purposes can also be expected to serve some formative function. This focus will be taken into account in this study.

2.1.2.3 *Summative assessment*

Summative assessment has been defined as "the assessments carried out at the end of an instructional unit or course of study for the purpose of giving grades or otherwise certifying student proficiency" (Shepard, 2006, p. 627). Put another way, summative assessment occurs when evidence is gathered in a planned and systematic way so that inferences can be made, based on reasoned evidence-based judgements about a student's overall achievement at a particular time, often with

the purpose of reporting this information to others (Harlen, 2005b, 2007; Harlen & James, 1996). This reporting can be to parents, to school boards and to government agencies. The information the report contains can be used for making decisions about resource allocation, student placement and student accreditation. It is its use for reporting and this kind of decision making that distinguishes it from other forms of assessment (Harlen, 2004). While summative assessment is necessary to provide a concise and accurate picture of learning and progress, Sutton (1992) cautions that it "flattens out the unique representation of the child as an individual, and sometimes produces an image more crude and blurred than we would like" (Sutton, 1992, pp. 3-4). There is considerable debate on what constitutes quality summative assessment in order to reduce the flattened representation that Sutton describes. Beginner teachers need to understand these debates and the many dimensions of summative assessment as well as its limitations if they are to implement it effectively in their classrooms.

Although formative and summative assessment have been presented in this chapter as two discrete purposes for assessment, there is considerable overlap. As concluded by Black (2013), "the formative and summative purposes of assessment can be so intertwined that they are mutually supportive rather than conflicting" (p. 176). Teachers can use assessment information formatively, summatively or for both purposes.

In the next section the various purposes of summative assessment are considered.

2.1.3 Focusing on the purposes of summative assessment

Summative assessment is used primarily to see whether and what learning has occurred for students at a particular time, usually at the end of a unit of work or the end of a course (Abell & Siegel, 2011). The judgements made about students' achievements, typically represented in the form of grades, percentages, marks and/or comments, are used for a variety of purposes within the school environment and external to the school (Harlen, 2005b). Within the school summative assessment results are used for ongoing teacher decision making and planning (e.g., streaming and class placements), and to report student progress to parents and to students themselves. External to the school, summative assessment results can be used for the accreditation of qualifications and for teacher and

school accountability (Harlen, 2005b). The following sections elaborate on these purposes in more detail: informing teacher decision making (Section 2.1.3.1), streaming classes and option placement (Section 2.1.3.2), reporting (Section 2.1.3.3), qualifications (Section 2.1.3.4), and accountability (Section 2.1.3.5).

2.1.3.1   Informing teacher decision making for teaching

A key purpose of assessment within schools is for tracking and analysis of student progress in order to inform teacher decision making regarding ongoing teaching plans (Abell & Siegel, 2011; Davies & Hill, 2009; Harlen, 2005b, 2007; Moreland, Jones, & Chambers, 2001; Weeden, Winter, & Broadfoot, 2002). In other words, summative assessment can be used for formative purposes by teachers. In a study involving 722 teachers in the USA, McMillan (2005) found that the teachers reported making instructional and assessment changes as a result of analysing summative assessment results. Similarly in New Zealand, Timperley and Parr have found that student achievement can be raised through this process (Timperley, 2009; Timperley & Parr, 2004). For example, when teachers provide in-class summative assessment tasks or examinations during the school year as practice for external examinations, the feedback from these can be used by the students in preparation for their later assessments, especially assessments which examine much of the same material. However, care is needed as such formative tasks can limit rather than enhance learning by narrowing the learning focus for the student and can lead to teaching to the test by the teacher (Crooks, 2011; Hume & Coll, 2008; McMillan, 2013). The use of school based summative assessment in preparation for external assessment in the NCEA context is discussed further in Section 2.3.2.

Data gathered from summative assessment can also be aggregated for groups of students and used for longer term teacher decision making such as planning for the following year as well as for the current class. This formative use of aggregated individual student summative assessment data has been debated in the literature. International research focusing on ongoing teacher decision making based on aggregated data argues that teachers are often unskilled in the use of aggregated data and tend to struggle to make best use of the data they collect (Datnow, Park, & Wohlstetter, 2007; Earl & Katz, 2006; Mandinach & Gummer, 2016; Pierce & Chick, 2010; Schildkamp & Kuiper, 2010). The knowledge

required to work with aggregated assessment data is specialised and requires technical expertise such as the knowledge of basic statistical testing, ways to graphically represent data sets, levels of significance and so on.

The complexities and subtleties of using aggregated summative assessment data necessitate specific instruction for preservice and inservice teachers so that they become aware and able to use it effectively. Researchers in the field of data literacy have identified the need for ongoing professional development and focused courses in ITE that are designed to help teachers develop the requisite expertise (Mandinach & Gummer, 2013; Schildkamp, Lai, & Earl, 2013; Unger, 2013). In their New Zealand work, Lai and McNaughton (2013) linked teachers' ability to use aggregated data effectively to their PCK. They found teachers with greater content specific knowledge about aspects of instruction were better equipped to interrogate the data appropriately and to make better decisions after they had analysed the data. Their findings indicate the use of summative assessment data has a discipline-specific element. This is recognised and investigated in this thesis study through a focus on the developing understandings of beginner teachers working to be secondary science teachers.

Some researchers note the limitations of using summative assessment for tracking and analysis of student progress in order to inform teacher decision making regarding ongoing plans. Harlen (2006, 2007), for example, argues that many summative assessment tasks do not match the principles of good formative assessment. Other researchers have found that the outcomes of particular assessment tasks can be analysed in different ways for use for different purposes (Black et al., 2003). This can be problematic (Black & Wiliam, 2003; Hill, 2000, 2006), as in general "the more purposes a single assessment aims to serve, the more each purpose will be compromised" (Pellegrino et al., 2001, p.37). Care must be taken in the decisions that stem from the use of data collected for one purpose being used for another, as this practice raises questions of validity and can produce unexpected impacts on students (Harlen, 2007). For example, poor results in a test on photosynthesis may lead a teacher to simplify the content of this unit for her next class, without realising the results were a reflection of a lack of engagement of the students, rather than the complexity of the content. The unexpected impact on students in the next class would be lessons pitched at too

low a level, restricting learning opportunities. As this example illustrates, it is important that teachers are aware of the limitations in using summative assessment to inform their decision making.

2.1.3.2   Streaming classes and option placement

A further purpose for summative assessment in schools is for streaming classes and option class placement, as well as informing the advice given by teachers to students when they move from one year level to the next (Black, Harrison, Hodgen, Marshall, & Serret, 2011). For example, a set of nominated summative assessment grades may be used to limit entrance to restricted classes such as Year 13 Chemistry or Physics, or be used to group the Chemistry students within a class by ability level. The use of summative assessment and standardised tests for streaming students is common practice, even though these assessment tasks may not have been designed for this purpose (Forgasz, 2010; Ireson, Clark, & Hallam, 2002). Research on streaming which is based on summative assessment has raised concerns over whether teachers actually do this in a dependable way (Black, 2013a; Hornby & Witte, 2014). It has been found that the grouping practices in New Zealand schools are not based on research evidence but on school managers who are not aware of the literature available on the topic (Hornby & Witte 2014). This suggests teachers, beginner teachers, need to be aware of the research evidence and issues relating to streaming and option selection in order to contribute more meaningfully to school decisions in these matters.

2.1.3.3   Reporting

The reporting of educational outcomes and learning achievement or progress to the students themselves and to their parents is a central purpose of summative assessment (Harlen, 2005b; Weeden, Winter, & Broadfoot, 2002). In line with current curricula (Ministry of Education, 2007, 2008) the reporting of student educational outcomes can include the level of development of student knowledge, skills and attitudes. Reporting on these outcomes can be by way of an overall mark, a letter grade, or an overall percentage that summarises these achievements. These may sometimes be accompanied by a narrative commentary from the classroom teacher (Brookhart, 2013a).

Whatever the means used for reporting student progress, the goal of reported summative assessment is "to give a trustworthy account of the extent to which

various learning outcomes have been achieved at a particular time, or to which learning has progressed over a specific time period" (Crooks, 2011, p. 72). The accuracy of reported grades, marks, percentages or narrative commentary depends on teachers' abilities to accurately summarise students' learning; it depends on the nature of teacher capability as assessors (Moss, 2013).

Teacher capability in assessment is a prerequisite to producing trustworthy reports of the extent students have achieved particular outcomes. However, questions have been raised in the literature regarding the trustworthiness of teacher report grades and comments. For example, one study, which focused on 12 New Zealand primary schools, found that teachers overwhelmingly wrote reports that were positively biased (Robinson & Timperley, 2000). This tendency to inflate grades and focus on positive comments was motivated by teachers' concerns about the esteem of students and families as well as their desire to increase community confidence in the school. Similarly it was found in a study which reviewed over 1000 school reports from New Zealand primary schools that 99% of comments made in reports were positive and that very few identified concerns and difficulties, with these few focused predominantly on behaviour rather than learning progress (Hattie & Peddie, 2003). Reporting overly positively on outcomes can mean that parents and students are not alerted to concerns about learning progress, and this may lead to feeling a false sense of security and/or to students experiencing difficulties and disillusionment at a later stage (Hattie & Peddie, 2003; Kofoed, 2009). Beginner teachers need to develop their capabilities for reporting summative assessment judgements in an honest and constructive manner.

2.1.3.4 Qualifications

Summative assessment results are also reported to agencies outside of schools for the purpose of awarding qualifications and for accountability measures. These are regarded as high stakes purposes of summative assessment because students' entrance to tertiary education or the workplace is often determined by their success in such summative assessments (Black, 2013b). The need for an accurate summary account of students' learning is highlighted when it contributes to high stakes qualifications. Summative assessment for qualifications needs to be based on procedures and criteria that guarantee comparability both within and between

schools. In New Zealand the grades derived from internal summative assessment based on standards are reported to NZQA and are used for the awarding of NCEA qualifications. These matters and NCEA specific summative assessment are discussed in more detail in Section 2.2.2.

2.1.3.5   Accountability

Ministry of Education officials and their counterparts in many countries view summative assessment as a tool useful for accountability purposes including, for example, for judging the effectiveness of teachers or schools, and for ranking schools (Gardner et al., 2010; Lane, 2013; Schneider, Egan, & Julian, 2013). However, the collection and use of aggregated individual student summative assessment data for this purpose is questioned on grounds of the reliability and validity of the evidence gathered for this particular purpose (Harlen, 2007; Thrupp & Easter, 2012; Thrupp, 2013). There is particular concern when school rankings based on aggregated summative assessment data from limited sources are published, leaving government officials and the general public to interpret the summary without important contextual knowledge (Thrupp, 2013).

The widespread use of summative assessment for accountability purposes can give the public an insight into the performance of students at school, but can have unintended consequences such as unfair public judgement about the standard of teaching in some schools leading to questions to do with school reputation and enrolment pressures. Therefore, as accountability in education becomes more prominent there is therefore an increasing requirement for teachers to develop robust summative assessment and data literacy skills (Athanases, Bennett, & Wahleithner, 2013; Jacobs, Gregory, Hoppey, & Yendol-Hoppey, 2009; OECD, 2013; Schildkamp & Ehren, 2013). There is also an increasing need for teachers to be able to engage with the wider public about what summative assessment results actually represent (Absolum et al., 2009; Chamberlain, 2013; Ministry of Education, 2011a).

In summary, summative assessment, which sums up student achievement at a particular time, can be undertaken for multiple purposes. Its results can be used both within and outside of schools. Within schools, summative assessment is under school control but sometimes also under the jurisdiction of external

agencies such as NZQA. It is conducted by teachers to measure student achievement in order to report learning progress to students and parents, as well as to inform ongoing decision making and class placement. Decisions are made for students based on both individual and aggregated summative assessment grades. Summative assessment is also used externally for the award of qualifications and for accountability purposes.

Beginner teachers need to gain an understanding of these various purposes of summative assessment in order to be able to develop and use summative assessment constructively with their classes. The challenge of ensuring quality in summative assessment has been hinted at in this section. Quality is discussed more fully in the next section.

2.1.4 Quality in summative assessment design and use

As a key component in quality teaching, quality in assessment has been identified as essential for raising student achievement by supporting learning (Absolum et al., 2009). Consequently, what constitutes quality in formative assessment practice has been the focus of much research over the last 20 years, and many lists of definitions and principles and practices have been published as a result (Absolum, 2006; Brown et al., 2008; Gardner, 2006; Harlen & James, 1996; Torrance & Pryor, 1998; Wiliam, 2006). Discussions to do with quality in summative assessment have a longer history. Reliability and validity are accepted as essential qualities in summative assessment, particularly summative assessment that is high stakes for students (Messick, 1994). Linked to validity, the principles of ethics and fairness have become more prominent in the literature recently as well as the associated impacts of assessment (Baartman, Bastiaens, Kirschner, & van der Vleuten, 2007; Cowie, 2015; Messick, 1994; Tierney, 2013; Tierney, Simon, & Charland, 2011). In countries such as New Zealand, where secondary teachers have the mandate to make decisions about assessments including designing and grading summative assessments that contribute to school-leaving qualifications, it is very important that they develop an understanding of the issues of quality within their summative assessment literacy. These quality issues encompass the principles of quality summative assessment (Section 2.1.4.1), quality task design (Section 2.1.4.2), and reliable and well-grounded individual

teacher judgement and moderation (Section 2.1.4.3). These issues are discussed in turn in this section.

2.1.4.1 Principles of quality summative assessment

To ensure quality in summative assessment, whether externally or teacher designed and implemented, it is generally agreed that it is important to consider both validity and reliability, and/or a combination of the two, which is captured in the concept of dependability (James, 1998). Additionally, the principles of ethics and fairness have been identified more recently as important to consider. In this section I outline how these principles are defined and discussed in relation to summative assessment.

Validity is a complex and multifaceted concept. Predictive validity (the extent to which performance in an assessment correlates to later performance on some criterion measure), consequential validity (the positive and negative, intended and unintended consequences of inferences based on assessment, discussed further in Section 2.1.5), face validity (the extent to what an assessment "looks like" it will assess what it is supposed to assess), and construct validity (the degree to which an assessment measures what it claims to be measuring) are just some of the facets that can be found in the literature. Each of these require consideration through the process of summative assessment design and use (Johnson, 2011).

The determinants of validity have been described as a linked chain of the following eight components of assessment:

> (1) Administration of assessment tasks to the student.
> (2) Scoring of the student's performances on the tasks.
> (3) Aggregation of the scores on individual tasks to produce one or more combined scores (total score or subscale scores).
> (4) Generalization from the particular tasks included in a combined score to the whole domain of similar tasks (the assessed domain).
> (5) Extrapolation from the assessed domain to a target domain containing all tasks relevant to the proposed interpretation.
> (6) Evaluation of the student's performance, forming judgements.
> (7) Decision on actions to be taken in light of the judgements.
> (8) Impact on the student and other participants arising from the assessment processes, interpretations, and decisions. (Crooks, Kane, & Cohen, 1996, p. 267–268).

Crooks et al. (1996) assert that each of the eight components is essential if valid inferences are to be made from either formative or summative assessment. They

explain that, for example, if in the administration of an assessment students get inappropriate help with an assessment task then validity will be undermined, even if the integrity of all other elements of the assessment chain are maintained. Similarly, if only a small sample of performance is used to assess a domain, generalisability will be limited. They propose that validity is limited by the weakest link in this chain. This means ensuring the validity of a summative assessment is a challenging task.

Validity has therefore been found to be a demanding aspect of summative assessment for teachers, with research showing that teachers display a lack of confidence and a weakness in their application of principles that guide validity (Black et al., 2010; Lovett & Sinclair, 2005). Additionally, it has been found that teachers' summative practices are not necessarily consistent with their beliefs about validity. Teachers' attention to validity can be undermined by external test regimes and by jurisdictions that link test results to school accountability. Research has shown that this could be redressed to some extent through professional development (Black et al., 2010; Stobart, 2001), nevertheless for most teachers the dominance of external pressures remain.

Reliability in summative assessment can be described as the degree to which an assessment task produces consistent and stable results over markers, time, occasions, items and tasks (Brookhart, 2003; Harlen, 2005a). Assessment task items need to be developed and used in ways that ensure assessment judgements will be reproduced if the same individuals are assessed again. This means, for example, that teachers and other assessment writers must design tasks and items so that a "sufficiency of information" (Smith, 2003, p. 23) is collected so that teachers have the information they need in order to make good decisions (McMillan, 2003). They must also construct marking schedules that are straightforward to use.

Research has shown that teacher judgement is not necessarily easy nor very consistent or reliable. Construct-irrelevant factors can influence the judgements of teachers when they mark student work, including gender, effort, behaviour, socio-economic status, ethnicity and personality traits (Bennett, Gottesman, Rock, & Cerullo, 1993; Brookhart, 2011; Brown, Moor, Silkstone, & Botton, 1998; Harlen,

2005a; Johnson, 2013; Klenowski & Wyatt-Smith, 2014; Mahuika, Berryman & Bishop, 2011). In a study of 200 teachers in the USA, McMillan and Nash (2000) focused on the factors that influenced grading, and proposed a model to explain how and why teachers decide on their practices for assessment and grading. It was found that a tension existed between teachers' internal beliefs and values and the realities of the classroom and other external factors, and this led to highly individualised and idiosyncratic assessment decision making and grading practices. This lack of consistency in assessment and grading practices is one factor that reduces the reliability of teacher summative assessments. Harlen (2007) argues that the reliability of teachers' judgements can be improved through a range of measures including the encouragement of professional collaboration and learning and the implementation of effective moderation processes. The contribution of moderation is discussed in detail later in this section.

Administration of tasks can also affect the stability of results, as can the condition of the person being assessed. For example, if students are fatigued, ill, unmotivated or anxious, they may not produce work consistent with their usual standard. Additionally, validity is likely to be reduced in this case, as assessment administration is a component in Crooks, Kane, and Cohen's (1996) linked chain that determines validity. This suggests that teachers' need to consider the aspects of their assessment practice that relate to both reliability and validity, if quality assessment is to result.

Dependability is a helpful concept to link between construct validity and reliability (Stobart, 2008). The trade-offs between these two elements are dependent on the purposes of the assessment. It is important not to over-compromise in either element. For example, it could be argued that reliability is more of a priority for a school-wide examination that is used to rank students for next year's class groupings than for an end of unit test on the process of photosynthesis which will be reported to students and parents. On the other hand, construct validity is important for the unit test on photosynthesis, as the results need to reflect the learning that has taken place for the student in that topic.

A number of conditions or influences on the dependability of summative assessment have been identified in the literature. From these, Harlen (2005a)

proposed five requirements for dependable summative assessment (i.e., assessment with construct validity protected and optimum reliability):

- Decisions about the domain of knowledge, skills and other attributes of learning to be assessed that are justified in terms of how learning takes place.
- A valid sample of student behaviour in the domain
- Criteria for judging the sample that are well matched to the goals of the work, of the curriculum and of the domain
- Procedures are in place for the reliable and unbiased application of the criteria
- Procedures for reporting and communicating with users of the assessment outcomes. (Harlen, 2005a, p. 263)

These requirements further demonstrate the breadth and complexity of summative assessment. They are of particular significance when classroom teachers are involved in the planning for assessment, gathering of evidence of learning, marking of the work and making judgements and decisions based on it.

The principles of ethics and fairness are connected to but distinct from validity, and also need to be applied to ensure quality summative assessment. In his review of literature focused on ethical and moral matters in teaching, Bullough (2011) noted that teaching is conceptualised as fundamentally and essentially a moral enterprise by most researchers. Even disciplinary knowledge has a moral dimension at its heart, and so shapes teachers' moral commitment and action (Elbaz, 1992). Green et al. (2007) define ethical behaviour as "acting based on one's judgment of an obligation—a duty by virtue of a relationship with a person, persons, or social institution" (Green, Johnson, Kim, & Pope, 2007, p. 1000). Their use of the term obligation suggests that assessment behaviour requires judgement in specific context, as obligations may not be static or generalisable and contexts may not be uniform. They also argue that "[e]thics is addressed in terms of obligations based on relationships so there is an affective component as well as a rational one" (p. 1000), so that obligations are felt rather than just followed.

New Zealand has a Code of Ethics for Certificated Teachers in New Zealand, and these must be applied to all aspects of teachers' work, including summative assessment. The professional interactions of teachers are governed by four fundamental principles:

Autonomy to treat people with rights that are to be honoured and defended
Justice to share power and prevent the abuse of power
Responsible care to do good and minimise harm to others
Truth to be honest with others and self.
(Education Council New Zealand, 2015, p. 1)

This broad code can be seen as being relationship-based and is a starting point when considering ethical behaviour for summative assessment.

The importance of an ethical framework within which to work has been highlighted with respect to assessment and its powerful impact (Gipps, 1994; Green et al., 2007; Popham, 2000; Shapira-Lishchinsky, 2011). Such a framework allows policy makers and teachers alike to draw upon it for assessment and other pedagogical decisions. Ethical guidelines for assessment such as *do no harm* and *avoid score pollution* have been suggested as principles which can help teachers make decisions, but given the nature of ethical decisions, uniform agreement on standards is not expected in every case (Green et al., 2007; Pope, Green, Johnson, & Mitchell, 2009).

It is evident in the literature that teachers are not necessarily equipped to make the ethical decisions involved in summative assessment that are required from them (Green et al., 2007; Pope et al., 2009). Because they work in an environment with sets of norms (e.g., interpersonal, professional, institutional, social conformity, self-protection), teachers face dilemmas when these norms come into conflict (Colnerud, 1997, 2006). For example, teachers may abandon the value of caring for students' wellbeing in assessment due to their need to submit to the views of senior leaders in their school. The organisation within a school may make dealing with these dilemmas more difficult for teachers as they struggle with their perceptions of institutional demands, the needs of students and other non-achievement factors (Colnerud, 2006; Pope et al., 2009).

Teachers have been found to have a limited understanding of the principles of ethical decision making, leading to varied practices when it comes to making decisions and judgements for such things as grading, testing, and assessing with students with special needs (Plake & Impara, 1997; Pope et al., 2009; Tierney et al., 2011). In practice, teacher judgement in summative assessment has been

found to be influenced strongly by teacher beliefs, knowledge, expectation and values as well as external policy (McMillan, 2003). Teachers have been shown to include factors unrelated to achievement when making judgements; factors such as their perception of student motivation, *pulling* for students by doing whatever will help students succeed, calculating grades without due consideration of the weighting of important assessment tasks, and assessing without defined success criteria. With respect to timing, for example, assessing performance assessments after a performance rather than during the performance can lead to inaccurate judgements being made (McMillan, 2003; Moss, 2013). Furthermore, teachers have been found to compromise standardised testing by a number of means such as teaching to the test, giving hints, changing time frames and so on (Brookhart, 1991; McMillan, 2003).

Tierney et al. (2011), in their study of Canadian secondary teachers working within a standards-based educational system, found that a better understanding of principles was needed for grades to accurately reflect students' learning, as the teachers were found to apply principles inconsistently. Many of the responses by teachers in this study suggested that their assessment practices were often guided by unexamined assumptions rather than by sound principles of assessment. More focus on assessment principles in teacher education to support the development of teacher judgement was recommended from this study.

The term fairness is sometimes used interchangeably with ethics. Fairness is a quality of assessment that is particularly problematic, given that fairness is accepted as a general principle of assessment, but it is interpreted in many different ways by teachers (Stobart, 2005; Tierney, 2013). Fairness is seen as more than a technical quality of assessment and extends to a concern for what is just within the context in which it is used. In this sense the power of those controlling the assessment, the social and cultural bias, and opportunity to have access to the curriculum, are all important factors that contribute to fairness (Cowie, 2015; Stobart, 2005).

Fairness would appear to be a universally agreed goal in assessment but many teachers struggle with the implementation of fairness in practice. Many equate fairness with using the same assessment task with all students, but on reflection

this apparent neutral approach may disadvantage some students; for example, those with reading difficulties. So although fairness is conceived as being equivalent to equality or consistency by some teachers (Tierney, 2014), many concede that equality does not imply fairness for all students, given their particular sets of circumstances and needs (Lyon, 2013b, 2013c; Tierney, 2014). Issues of fairness, such as opportunity to learn and demonstrate learning, language of instruction and assessment, student engagement, classroom environment, and teachers' views on the tension between equal and equitable treatment of students, all contribute to the fairness of assessments (Moss, Pullin, Gee, Haertel, & Young, 2008; Sluijsmans & Struyven, 2014).

Although there has been only some degree of consideration in the literature on equity and ethics with respect to assessment, more recently there has been a shift towards understanding that equity, rather than equality and consistency, leading to fairer assessment especially in terms of supporting learning (Lyon, 2013b, 2013c; Moss et al., 2008; Tierney, 2013). This is of particular interest in school populations which are very diverse in terms of culture, language and social background. A move to more differentiated assessment by using pedagogy that is responsive to students and meets their needs is seen as more enabling for students (Mahuika et al., 2011; Moss et al., 2008; Wyatt-Smith, Klenowski, & Colbert, 2014).

In summary, an awareness of the principles of good assessment, including ethics and fairness, for classroom-based summative assessment is necessary for all teachers. Given the stakes involved and the complexity of assuring quality in summative assessment, it follows then that these requirements should be a learning focus for beginner teachers so that their assessment practices are valid, reliable, dependable, ethical and fair. This thesis investigates how beginner teachers develop an understanding of the principles of quality summative assessment.

2.1.4.2 *Task design in summative assessment*
Whether summative assessment is carried out under the jurisdiction of a classroom teacher or an external agency, the accuracy of the assessment findings is dependent on the quality of the assessment task and process as well as the

capability of the assessor (Brown et al., 2014; Moss, 2013). Well-designed tasks allow teachers to make judgements and decisions based on better quality information. For quality assessment tasks, teachers need to be able to apply the principles discussed above: validity, reliability, ethics and fairness. They require considerable curriculum and assessment knowledge so that the strategy or task they design allows the collection of evidence of learning that is matched to their achievement objectives and is appropriate for their students (Black, 2013a; Moss, 2013). However, many teachers have had little or no training in assessment task design (Lyon, 2013c; Whitehead, 2007).

In her review of research on classroom summative assessment, Brookhart (2013b) found that teachers were inadequately prepared to design assessment tasks, even though some were confident in their ability to do so. It has also been found, based on content analysis research, that many teacher-designed tasks are pitched at low cognitive levels, and teachers adapt tasks rather than develop their own (Shepard et al., 2012). A New Zealand study focused on the preparation of New Zealand secondary teachers found that they lacked confidence in assessment design (Anthony & Kane, 2008). These findings are of relevance for this study because even beginner secondary science teachers have the responsibility to design or choose relevant summative assessment tasks for their students. If they are teaching students in senior secondary school classes they will also be responsible for assessing internal standards for NCEA.

A wide range of assessment strategies and tasks can be used to support students to produce evidence of their learning in science classrooms. These include traditional tasks such as tests, examinations, laboratory reports, science investigations, essays and speeches as well as more innovative tasks such as designing a website or game, group problem solving and so on. In some jurisdictions the use of externally supplied tests or examinations is a key form of summative assessment (Brookhart, 2013b; Morrison & Tang, 2002; Pong & Chow, 2002). However, in many countries teachers have at least some control or choice over the form of summative assessment used within their classroom and sometimes as part of exit qualifications. The use of authentic assessment and naturally occurring evidence is becoming increasingly important and accepted around the world, especially by those working towards aligning their educational outcomes to twenty-first century

skills and competencies (Koh, 2014). Koh (2014) emphasises the necessity for teachers to have high levels of assessment literacy if authentic assessment is to be used to promote student learning in high-accountability contexts.

Current trends indicate a move towards the use of summative assessment tasks which are more authentic and towards tasks that produce performance-based evidence. These forms of tasks are seen to be more meaningful to students, and hence help students to learn and to express what they know more effectively (Gardner et al., 2011; Hickey & Anderson, 2007; Hipkins, 2012; Koh, 2014; Shepard, 2000; Venville et al., 2008). They can also be more aligned with what is required for life and work in the fast changing world, including "the skills, knowledge, attitudes and characteristics of self-directed and collaborative learning that are increasingly important for our global economy and fast changing world" (Kozma, 2009, p. 1). Teachers are being called to assess students in new ways that address twenty-first century needs (Scardamalia, Bransford, Kozma, & Quellmalz, 2012). Two examples of innovative assessment tasks in science follow:

(i) extended experimental investigations used in Queensland (Australia) included tasks which demanded that students were in full control of the investigation (Darling-Hammond, Herman, Pellegrino et al., 2013). Students had to articulate hypotheses, undertake experimental design (show evidence of modification or own design), produce primary/secondary data, execute experiments, analyse any data, discuss the outcomes, evaluate and justify conclusions, and present a scientific report. Similar investigative assessments are used in New Zealand for senior science achievement standards, and these require students to control the whole inquiry and report on its findings. These investigations take place over a significant period of time and allow for teacher interaction and some guidance. Teachers assess aspects of the investigation from the planning stages, and throughout the process, as well as the final report (NZQA, 2017b).

 (ii) In a study reported by Gale et al. (2016) a simulation-based performance assessment was used with 13-year-old students in the USA to assess their understanding of energy, forces and motion. Using LEGO Mindstorms kits students worked as traffic engineers to understand why heavily loaded trucks

leaving a factory were involved in accidents. Then as mechanical engineers they designed an automatic brake for the trucks. The conversations teachers had with students during the performance assessment tasks meant the students were engaged in meaningful discussion related to their understanding of these science concepts. Students were assessed during the course of the unit via their engineering notebooks, their engagement with problem-solving scenarios, and later with externally supplied multi-choice tests. The simulation based performance assessment yielded useful insights about student learning as well as information to inform future curriculum revisions.

Summative assessment tasks, such as those illustrated above, allow assessment of student achievement in more meaningful and authentic ways. They illustrate that summative assessment tasks can take a number of forms and can involve students in activities more closely aligned with teaching and learning activities and with actual science practices. Overall, the range of assessment formats and strategies available to teachers is expanding, and so teachers need assessment literacy in order to make best decisions that lead to quality assessment in their classrooms. Unfortunately, research suggests that on the whole teachers' abilities to design quality tasks are low, which increases the risk of variable quality assessment. While the use of authentic tasks raises interesting possibilities for teachers, they need support to make informed use of these possibilities.

2.1.4.3   *Individual teacher judgement for summative assessment purposes*
Quality summative assessment involves teachers making substantiated judgements about learners' achievement, based on their attention to the evidence of learning they have gathered, and appraising this against criteria or standards, or some other reference framework (Sadler, 1998). Students' performance can be compared through the use of norm-, criterion- or self-referenced assessments. Norm-referenced assessments allows a student's performance to be compared relative to a peer group, rather than to an absolute performance level or standard. These assessments tend to be developed to spread out the performance of students to enhance showing relative performance. Norm-referenced assessment is widely considered to be unfair given the complexity of contextual factors, questions over the need for standardisation and the necessity to compromise what is assessed (Gipps, 1994). As an alternative, criterion-referenced assessment, or standards-

based assessment, compares a student's performance against established criteria or standards. This allows an estimate of the absolute level of student performance to be made as the evidence of their learning is compared against the standard. This form of assessment requires very clear understanding and application of criteria or the standard during marking. A third method of judging student progress is self-referenced or ipsative assessment. In this instance, a student's performance is compared to their own previous performance. This type of comparison enables individual progression to be assessed and monitored (Brookhart, 2013a). Quality teacher judgement is required no matter the reference framework of the assessment used.

Classroom teachers are arguably in an ideal position to make judgements about students' learning, as they can sample a wider range of students' work than an external assessor or single assessment is able to, which enhances both validity and reliability of the information and the judgement (Mansell, James, & Assessment Reform Group, 2009). Classroom teachers are also able to integrate the many elements of performance that are involved in authentic assessment tasks, students' engagement in practical investigations, and aspects of student creativity and creative reasoning, in ways that external examinations do not allow (Black, 2013b). In a review of research on summative assessment practices, Moss (2013) found that the literature is divided on the ability of teachers to make quality judgements about student learning, with some researchers finding that classroom teachers were better placed to assess their own students given their knowledge of them, but others finding that teachers' judgements of performance was compromised by their inability to distinguish between student attainment and other traits such as behaviour and motivation. The idiosyncratic nature of teacher conceptions and background experiences (discussed in more detail in Section 3.1.3.2) and the lack of deep knowledge of criteria and standards goes some way to explaining the inconsistencies evident in teacher judgements.

Bonner and Chen (2009), in their investigation into the developing grading practices of preservice teachers in City University of New York, found differences between elementary and secondary teachers. Their research confirmed that with regard to making assessment judgements preservice teachers were very concerned about fairness and not being seen as too lenient in their grading. Overall they

found the teachers' thinking about nonstandard grading practices to be multidimensional, affected by a range of concerns such as the need to be academically enabling and benefit the learner group, classroom management, laxity in grading, and their thoughts about alternate assessment formats. Secondary teachers tended to be more likely to view grades as a way to punish students or manage student behaviour, and this related to more of their number being likely to align their practice with traditional teaching approaches. However, many teachers from both sectors were found to have adopted a constructivist teaching philosophy, and these teachers tended to be strong in their support of enabling approaches, and the use of alternative assessments. By completing a university course focused on assessment, preservice teachers' support of nonstandard grading practices became more moderate, and this was suggested to be a result of their learning about fairness and validity with respect to grading and grade interpretation.

In a commentary of the New Zealand qualifications systems, Black (2001) noted the challenges that teachers face in making judgements for internal assessments for the school leaving qualification, NCEA:

> Those aims of education which are properly reserved for teacher assessment are those where holistic and complex judgments are called for, and for which the background of experience that is provided by long traditions of external testing is not available, so that it is inescapable that some of the most complex expertise is called for in areas where teachers' experience is most weak. (Black, 2001, Section 4.1)

The judgements that teachers are called on to make require considerable expertise, and teachers need to build this up over time. Black (2001) comments on the importance of providing teachers with real examples of student work along with commentaries so that they are able to build a deeper understanding of where the 'standard' resides. Consistency of and between teacher judgements is a major concern when classroom summative assessment is used for the awarding of high stakes qualifications (Klenowski & Wyatt-Smith, 2014). Therefore involvement in social moderation is an essential mechanism in helping teachers understand criteria at a deep level, and to ensure some consistency in their judgements.

The process of moderation is an important aspect of quality assurance for teacher summative assessment, both internal and external to a school (Gipps, 1994; Harlen, 2005a; Maxwell, 2007). As noted by Wyatt-Smith and Gunn (2009, p. 94), "a way forward is to recognize that teacher judgment, in conjunction with clearly specified standards and opportunities for moderation, are a linchpin of a robust assessment culture in schooling". Put simply, moderation is a social process for ensuring consistency in teacher judgement of student work. Moderation involves teachers interacting with one another, sharing knowledge relating to criteria, and marking student work together by referring to the standards and exemplars, in order to gain a common understanding on which to base their assessment judgements. The overall purpose of this process is to produce valid and reliable teacher judgements that are consistent both with the standards and with one another (Klenowski & Wyatt-Smith, 2014).

The knowledge base of teachers, both explicit and tacit, and their representation of standards are factors that influence the moderation process, as it is not a clear and linear process (Adie, 2011; Adie, Klenowski, & Wyatt-Smith, 2012; Wyatt-Smith, Klenowski, & Gunn, 2010). Teachers' interpretations of standards are shaped by their own experiences of being assessed and by their contexts (Sadler, 1987, 2009). As teachers interact when looking at samples of student work and related standards in order to make judgements, they make their work visible to other teachers and negotiate understandings of the meaning of standards within an assessment community. Through this process their own explicit and tacit knowledge interplays with the group's knowledge and common understandings develop. Moderation meetings provide opportunities for negotiation and can lead to the development of new knowledge. The interaction involved in moderation has been found to deepen collegiality, build confidence, develop content and pedagogical knowledge and provide professional learning opportunities (Barrett, 2008; Gipps, 1994; Hipkins & Robertson, 2011).

As well as school-based moderation processes, it is important to have a system to check the consistency of judgements that teachers are making across a state or country, especially when these judgements contribute to the award of qualifications (Klenowski & Wyatt-Smith, 2014). The existence of a governmental statutory body responsible for the oversight and quality assurance

of formal summative assessment is a feature in most countries in the world, particularly with respect to exit qualifications from secondary schools, and for state or national testing for other age groups. In New Zealand, schools accredited to provide qualifications must develop and use internal processes that check for quality tasks matching the standard and achievement criteria, and consistency of judgement. As well as this, a national external moderation system is in place to enhance consistency of judgement within and across NCEA assessments. This system provides robust review of and feedback on assessment materials and the marking of student work. It is designed to ensure the public can trust the qualifications awarded (NZQA, 2016a). More about New Zealand NCEA moderation is discussed in Section 2.2.2.3.

In summary, quality is of central concern in summative assessment. The principles of good assessment — validity and reliability (sometimes combined as dependability), ethics and fairness, must be understood and applied by teachers in the same way as other assessors, if assessment results are to be trustworthy and have any standing within the wider community. Teachers also need to understand the elements of quality task design and the factors involved in making quality judgements about student achievement, including the process of moderation. Because of their responsibility for assessment, beginner teachers in New Zealand need to develop a thorough understanding of the principles and practices that lead to quality summative assessment.

2.1.5   The consequences of summative assessment for teachers and students

Summative assessment has been found to have an impact on students and teachers, and on the learning and teaching that occurs in classrooms (Carr et al., 2005; Harlen, 2004; Harlen & Deakin Crick, 2002; Hipkins & Neill, 2006; Hutchings, 2015; Keddie, 2016; Pratt, 2016; Shepard, 2000). The consequences of summative assessment for students include the qualifications they are able to gain as well as their level of motivation and self-perception as learners. Summative assessment influences teachers' pedagogical decision making and action as well as leading to personal responses, especially where accountability plays a role. The consequences of summative assessment can be long lasting and hard to alter, so care needs to be taken that its role is to "embrace the whole work of learning and it has to relate to the whole work of teaching, and then to assess the effectiveness

of this combined work" (Black & Atkin, 1996, p. 116). In this section, literature related to the consequences that summative assessment can have on students and their learning (Section 2.1.5.1), and teachers and teaching (Section 2.1.5.2) is reviewed.

2.1.5.1 *Consequences for students and their learning*

The pressures and constraints generated by summative assessments, such as national examinations or tests, has been well documented (Crooks, 1988). Research shows that such summative assessment can affect students' motivation, future achievement, attitude, and so on (Crooks, 1988; Hipkins, 2010; Hipkins & Neill, 2006; Moss, 2013). These consequences can be very visible and public, and hard to undo when, for example, they influence student subject selection and placement in classes. Some consequences are less visible and involve the affective domain. In particular, assessment has been found to affect student motivation and more broadly student self-regulation, both positively and negatively (Brookhart, 2013c; Keddie, 2016).

An important mediating factor for the consequences of summative assessment on students is the involvement of their emotions. Emotions are complex and a multifaceted phenomenon and can be triggered by the learning environment, assessment design, performance expectations, assessment feedback and consequences (Pekrun & Linnenbrink-Garcia, 2014; Vogl & Pekrun, 2016; Zeidner, 1995). Researchers have found a range of emotions to be linked to assessment for students and that these influence their motivation, activation of cognitive resources, learning behaviours and assessment outcomes (Brown & Wang, 2013; Carless & Lam, 2014; Harris & Brown, 2009; Pekrun et al., 2004; Spangler, Pekrun, Kramer, & Hofmann, 2002). Most research focuses on test anxiety and its effects on performance and validity (Brown, Irving, Peterson, & Hirschfeld, 2009; Pekrun et al., 2004; Pekrun, Elliot, & Maier, 2006; Vogl & Pekrun, 2016). An extensive body of research demonstrates the detrimental consequences of anxiety for most students. However, research also draws attention to positive learning environments in which the challenge of carefully designed assessment tasks can be enjoyed by students, thus increasing the potential of such tasks to assess students' actual capabilities (Vogl & Pekrun, 2016).

In her review of the evidence of how the process of summative assessment affected students, teachers and curriculum, Harlen (2004) synthesised 23 relevant studies. Evidence suggested that students benefitted when teachers helped them to understand assessment criteria and expectations, and when teachers' assessment practices included processes and explanations. The synthesis found that students' motivation has been shown to be affected by teachers' orientations in summative assessment: using grades for rewards or punishments was demotivating, whereas providing non-judgemental feedback (for example, in portfolio style summative assessment) was motivating.

Other scholars have reviewed the relationship between summative assessment and motivation and student outcomes and have found there to be a range of factors involved, so the link between assessment and motivation is complex rather than a simple cause and effect relationship (Brookhart, 1997; Harlen & Deakin Crick, 2003). More recently self-regulation theory has grown to subsume motivation theory, and it has been suggested that the self-regulation of learning includes both formative and summative assessment, with the student being at the centre of the process (Brookhart, 2013c). As an example of student self-regulation engagement, a study of 3790 secondary students from 20 schools in New Zealand found that students' motivation orientations for NCEA included *doing my best* and *doing just enough*, and were strongly related to actual achievement (Meyer, Weir, McClure, Walkey, & McKenzie, 2007, 2009). An interesting aspect of these studies, illustrating complexity, was that students' orientations were seen to be correlated to gender, ethnicity and the socio-economic level of the school community.

The way students view themselves and their success as students can also be a consequence of their experiences with summative assessment. For example, in their review Harlen and Deakin Crick (2002) found that following the introduction of national tests in England, there was a correlation between student self-esteem and achievement with high achieving students being more likely to have higher self-esteem than their lower achieving classmates. In her study of Years 5 and 6 students in England, Keddie (2016) found that students were affected by the neoliberal practices in their school, and viewed their worth and value as being closely linked to their individual performance in summative assessments. Keddie's study highlighted the narrow focus students had when they

considered what counted as a good student and a good citizen. Similarly, in his study of secondary school students in Australia, Thompson (2010) found that the students had a limited vision of a good student as one who performed well in assessment, behaved well, and displayed values acceptable with dominant expectations. However, students limiting their view of success to their assessment grades is in contradiction to the aims of school curricula calling for independent life-long learners, capable of creative and divergent thought, able to work collaboratively (Hutchings, 2015; Thompson, 2010).

Summative assessment has been found to dominate in classrooms and distort the curriculum experienced by students, giving high priority to some aspects of the curriculum and narrowing its potential (Carr et al., 2005; Harlen, 2005a). However, the consequences of teacher-controlled assessment for students have been found to vary depending on how teachers view their role as assessors, that is, whether teachers have focused on improving the quality of students' learning or on maximising students' marks (Bullock, Bishop, Martin, & Reid, 2002; Yung, 2002). Assessment can provide a positive impact on student learning when learning rather than performance is emphasised by teachers, and when testing and instruction are integrated so that feedback and remediation are provided (Carr et al., 2005; Harlen, 2005a). Bullock et al. (2002) found that in high stakes contexts the pressures of achieving good grades mitigated against students developing less tangible skills such as critical thinking, creativity and independent learning. However, in low stakes contexts teachers were shown to clarify students' understandings of learning outcomes, and together with an extended range of assessment tools this is more likely to result in an extended range of learning experiences for students. Students' shared understandings of learning outcomes and assessment criteria was seen to be crucial if an optimal level of high order thinking was a goal and was embodied in summative assessment.

2.1.5.2 *Consequences for teachers and teaching*

Summative assessment has been shown to have consequences for teachers and their teaching. It is well documented that school assessment regimes affect the way teachers present the curriculum to students and the way teachers operate in their classrooms (Carr et al., 2005; Donovan & Bransford, 2005; Gipps et al., 1995; Hipkins & Neill, 2006; Hume, 2006; Levin & He, 2008; Popham, 2009;

43

Smith, 2010; Thrupp & Easter, 2012). Yung (2002) found that some teachers adopt a passive role regarding national assessment policy interpretation and implementation and just do what it says, whereas others have the professional confidence to adopt a more critical stance when interpreting policy requirements, demonstrating a more proactive response to policy and viewing it as an opportunity to reconstruct their pedagogy. Yung argues that a teacher's confidence and professional consciousness, whether this focuses on student interests or teacher self-interest, has a strong bearing on the stance on assessment decision making. Similarly, in the area of formative assessment researchers have also found this to be the case, with teachers' beliefs being found to be influential in regard to their response to policy directions (Dixon, Hawe, & Parr, 2011; Pedder & James, 2012). Therefore in the classroom both formative and summative assessment are affected by teachers and generate consequences for teachers.

Summative assessment can affect teachers' pedagogy positively. When teachers use assessment strategies that require deeper thinking from students, as a consequence they are more likely to engage in teaching practices that extend students' learning experiences (McMillan, 2001). Black et al. (2011) described a study which explored how teachers learnt about summative assessment through the use of five strategies. The findings of this longitudinal study demonstrated that the use of these strategies enabled teachers to enhance their practice, producing positive outcomes for students, parents and teaching colleagues. These included better designed tasks that took account of uniformity and diversity, increased understanding of validity in summative assessment, growing teacher confidence and collaborative practice.

There is evidence, however, that the positive effect of assessment on pedagogy is predicated on teachers having relevant PCK (Black, 2003). For example, it is with a thorough understanding of the science concepts being assessed, of common misconceptions, of learning progressions for the concept, of effective teaching strategies that science teachers are better equipped to assess learning in a way that will provide more informative formative and summative assessment data to the teacher (Corrigan, Buntting, Jones, & Gunstone, 2013; Schneider & Plasman, 2011).

However, some research points to negative consequences of assessment for teachers when control of curriculum is restricted. High stakes summative assessment has been found to influence teacher curriculum implementation and distort the intentions of a curriculum if it dominates by narrowing the curriculum (Black, 2013b; Hutchings, 2015; Pratt, 2016; Thrupp & Easter, 2012). Teachers can feel obliged to prepare students for assessments, especially when school-set targets for student success are prioritised (Stobart, 2008). In this sense the assessment of student learning is experienced by teachers as an assessment of their performance (Rutherford, 2003). The use of student summative assessment results as a measure of teacher effectiveness can weigh heavily on teachers, with a number of studies indicating that when this happens it tends to create tensions in teachers' practice that distort their approach to teaching and to assessment (Ball, 2003; Pratt, 2016). For example, Pratt (2016) found that teachers adjusted grades down at the beginning of the year and commented that "teachers' decisions about where children are in relation to assessment levels is a complex amalgam of their desire to act in a manner which is fair to pupils, the need to grow their professional capital and the question of what is negotiable and what is not" (p. 11).

As with students, teachers' emotions are an important consideration within their work as assessors. The emotional practice of teaching has been well documented with a number of researchers linking emotion with teaching and with teacher cognition, as well as with the social cultural and historical contexts in which teachers work (Fried, Mansfield, & Dobozy, 2015; Hargreaves, 1998; Nias, 1989; Zembylas, 2004). Research shows that both positive and negative emotions may be strongly experienced by teachers as they take responsibility for students' progress (Näring, Briët, & Brouwers, 2006; Nias, 1989). Teachers may experience positive feelings when acting in a way consistent with their beliefs, and negative feelings, including vulnerability, if being judged by themselves or others or when they are acting in ways that conflict with their beliefs and values e.g., when policy dictates a loss of their autonomy (Bell, 2011; Kelchtermans, 1996; Lasky, 2005). A common emotion discussed in recent literature is anger or frustration stemming from either a lack of institutional support or students' indifference or misbehaviour (Fried, Mansfield, & Dobozy, 2015). This is linked to a sense of

powerlessness and can lead to anxiety, especially when the teachers consider that student achievement may be used to judge their effectiveness (Nias, 1996; Shapiro, 2010). Given that emotion and cognition are fully integrated (Fried, Mansfield, & Dobozy, 2015; Neville, 2013; Vygotsky, 1978) it follows that teachers' assessment decisions are directed at least to some extent by their emotions, although substantial research has not been completed in this field.

In summary, this section has described some of the consequences of summative assessment for students and teachers which are very real, long lasting and need to be acknowledged. Teachers need to develop an awareness of the effects of assessment on their students and on themselves in order to make decisions that lead to positive learning outcomes and trustworthy summative judgements and grades.

The following section reviews literature focused on summative assessment in the New Zealand context, including that used for NCEA.

2.2 Summative assessment in the New Zealand context

The New Zealand school system, and the policy and practices particularly relevant for summative assessment, provide the context for this study. In this section I provide background information and review research literature that specifically focuses on summative assessment within the New Zealand secondary school context. Firstly, a brief outline of the secondary school system is provided, together with a description of policy documents that influence summative assessment and direct teachers' practice (Section 2.2.1). This outline provides the contextual knowledge base for the study reported in this thesis. Subsequent sections focus on the NCEA qualification system (Section 2.2.2), teacher and student experiences of NCEA (Sections 2.2.2.1 and 2.2.2.2), and quality assurance mechanisms of the NCEA qualification (Section 2.2.2.3).

2.2.1 The New Zealand school structure and system
New Zealand has a highly devolved school system in which schools are self-managed and self-governed under a Board of Trustees (BOT), comprised of elected parents and staff from the school community. The BOT is responsible for all management and governance decisions and is charged by the New Zealand Ministry of Education to meet the requirements of the National Education

46

Guidelines (NEGs) which establish a common direction for state education in the country. The NEGs include National Educational Goals, foundation curriculum policy statements, national curriculum statements, national standards (for Years 1–8 students), and the National Administration Goals (NAGs) (Ministry of Education, 2009). Schools are audited by the Education Review Office (ERO), an independent review agency, to try to ensure they meet these requirements.

The range of school structures in New Zealand are shown in Figure 1 below. In this figure the ages, year levels and qualification levels are shown in the left hand side of the diagram. It can be seen that formal qualifications are normally gained by students in their final three years of secondary schooling.
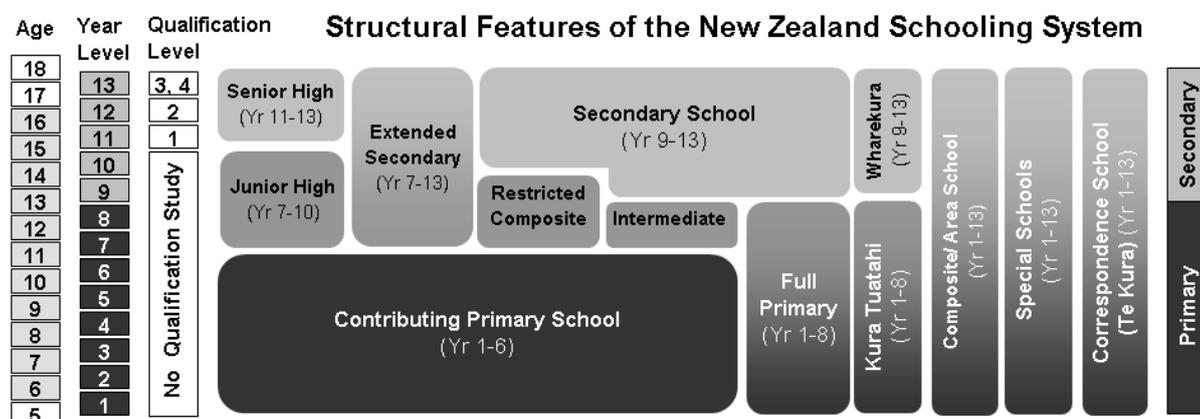
Figure 1.   The New Zealand education system

Reproduced from Ministry of Education (2011b, p. 16)

The regulatory framework provided by the Ministry of Education includes requirements associated with summative assessment. These include "the establishment of clear learning objectives, monitoring student performance against those objectives" (NEG 6) (Ministry of Education, 2015, para. 8), and "access for students to a nationally and internationally recognised qualifications system to encourage a high level of participation in post-school education in New Zealand" (NEG 8) (Ministry of Education, 2015, para. 11). Schools must gather comprehensive information so that student progress and achievement can be evaluated. On the basis of this information, teachers must identify groups of students and aspects of the curriculum that require particular attention, and develop and implement teaching strategies to address these needs (Clarke, Timperley, & Hattie, 2003; Davies & Hill, 2009). Schools must also maintain a

programme of self-review, which includes a review of information on students' achievement (Ministry of Education, 2015). The New Zealand Curriculum (NZC) states that effective assessment benefits students, involves students, supports teaching and learning goals, is planned and communicated, is suited to purpose, and is valid and fair (Ministry of Education, 2007). This suggests an improvement-focused approach to all assessment, including summative assessment.

In 2009 an important policy advice paper, *Directions for Assessment in New Zealand (DANZ)* (Absolum et al., 2009), was produced. This built on an earlier strategic review of assessment strategy (the National Assessment Strategy) in 1999 and promoted the idea of teachers and students being assessment-capable, defined as being "able and motivated to access, interpret and use information from quality assessments in ways that affirm or further learning" (Absolum et al., 2009, p. 6). *DANZ* and a resulting Ministry of Education position paper on assessment for the schooling sector (Ministry of Education, 2011a) raised issues related to teacher education, noting that support was required in order to facilitate an active assessment community. The 2011 Ministry of Education position paper emphasised the need for a high level of responsiveness by schools and teachers to the unique learning contexts that present themselves in individual schools (Ministry of Education, 2011a). Culturally responsive pedagogies (which include assessment practices) are prioritised in the New Zealand context, particularly with respect to reducing disparities between higher achieving Asian and European students and generally on lower achieving Māori and Pasifika students (Mahuika et al., 2011). There has been a call for teachers to include assessment practices that meet the needs of Māori, Pasifika and students with special education needs as a priority (Mahuika et al., 2011; Ministry of Education, 2011a). This demand adds another dimension to the assessment literacy required by teachers in all levels of schooling in New Zealand.

In the junior levels of secondary schooling (Years 9 and 10), classroom teachers are responsible for the summative assessment used in their classes. Although there are some standardised tests available their use is not mandated and the results of these standardised tests are not necessarily used by classroom teachers to inform their teaching or for reporting (Heron et al., 2001). Most Year 9 and 10 science

assessments are in the form of tests or assignments and are conducted in class, although some schools formalise the end-of-year assessments by setting examinations. These are designed, delivered and marked by classroom teachers, often in ways that mirror NCEA assessment grading. In their study of New Zealand secondary teachers, Irving, Harris, and Peterson (2011) found that NCEA programmes dominated teacher thinking about assessment and that it had obvious washback effects into Years 9–10. A national survey carried out by the New Zealand Council for Educational Research (NZCER) in 2012 reported unease from respondents (secondary school principals, teachers and BOT representatives) who thought that assessment was driving the curriculum across their schools, including for students in Years 9 and 10 (Wylie, 2013). This perception had increased from what was found in 2009. The 2015 survey also found NCEA was considered to impact on Years 9 and 10, with some teachers giving their classes examinations that reflected those used in NCEA thereby showing a trickle-down or washback effect of the qualification system. Most teachers reported feeling under pressure to improve their students' results. The system of assessment for Years 11–13 students relates predominately to NCEA.

2.2.2 The NCEA qualification system in detail

For senior secondary students (Years 11–13), summative assessment is focused primarily on qualifications available on the New Zealand Qualifications Framework, particularly the NCEA at three levels, Levels 1, 2 and 3. NCEA must be offered in every public and integrated secondary school in the country. The NCEA qualifications are gained through a standards-based assessment system which was introduced in New Zealand in 2002 to replace a norm-referenced examination-based system. Two sorts of standards can be used by teachers and students to generate credits for NCEA: achievement standards (ASs) which are based on the *New Zealand Curriculum*, where standards at levels 1, 2 and 3 equate to *New Zealand Curriculum* levels 6, 7 and 8 (usually corresponding to students' final three years at secondary school); and unit standards (USs), which are used to assess work place and industry related learning. Each AS and US has an associated credit value which relates to the amount of learning needed to cover the material assessed: one credit is equal to ten notional hours of learning (NZQA, 2001). Students are able to work in a modular fashion to build an NCEA award,

and the level at which they are working does not have to be linked to their chronological age (Hipkins, Vaughan, Beals, Ferral, & Gardiner, 2005). NCEA awards can be gained through a combination of standards. For example, for a NCEA Level 1 Science class, teachers could choose a selection of standards from a range of 16 core Science ASs, each worth 4 credits, as well as 5 ASs for each of Biology, Chemistry and Physics. From this group of over 30 standards most teachers would choose up to 6 standards in total for a class or an individual student, with a mix of internally and externally assessed standards.

An NCEA award is gained by the accumulation of credits. A total of 80 credits with 60 credits at or above the level of the certificate is required for the award of each of the levels of NCEA. Usually students are assessed against standards that contribute approximately 90–120 credits per year, or approximately 20 credits per subject, for a full year programme, but this is at the discretion of the school. More than half of the ASs and all of the USs are internally assessed, that is, tasks are chosen or designed, used, marked and checked within a school by classroom teachers and the results are then reported to NZQA. For each internally assessed standard teachers are provided with the standard, which outlines assessment criteria, guidelines for assessment, exemplars of assessment tasks and of marked student work. Teachers have the freedom to adapt existing tasks or develop new assessment tasks that meet the requirements of the standard. They are expected to work with at least one other teacher to critique assessment materials before they are used and internally moderate the marking of student work when tasks are completed. Consequently, social moderation of teachers' judgements has become mandated practice for teachers in New Zealand senior secondary schools since the introduction of NCEA in 2002 (NZQA, 2015a). A further quality assurance process is administered by NZQA in which at least 10% of internally assessed standards are externally moderated. Schools are asked to provide random samples of assessment materials for standards nominated by NZQA, so these can be checked for quality by a moderator external to the school. Additionally, a small number of standards (up to three for any one subject at a particular level) are externally assessed by end of year external examination and this process is managed externally by NZQA.

Each AS can be graded at four levels of competence: Excellence, Merit, Achieved, or Not Achieved. Most USs are simply graded on an Achieved/Not Achieved basis. If students are awarded an Achieved grade or higher for any standard, they are awarded the credits associated with the standard. NCEA awards can be endorsed with Merit or Excellence if students get 50 credits or more at that level of performance, and individual subjects can also be endorsed within certain guidelines. Endorsements to NCEA qualifications were introduced in 2008 (at certificate level) and 2010 (at subject level). In her review of NCEA, Hipkins (2013) found that the endorsements provided in NCEA were seen as a positive move by teachers and school leaders.

One further distinctive feature of the NCEA qualification system is the provision for fair assessment (NZQA, 2001). Teachers are charged with giving all students the opportunity to provide evidence of the learning, and this may at times involve adapting assessment tasks to enable students who have special requirements to be able to complete the tasks. NZQA provides approval for special assessment conditions so that "entitled candidates can demonstrate their knowledge, skills and understanding, without providing unfair advantage over other candidates" (NZQA, 2016b). Special assessment conditions apply to both internally and externally assessed USs and ASs.

As can be inferred from this description this assessment system requires summative assessment literacy from all secondary classroom teachers who teach students in Years 11–13. All the more so, as reported in a recent OECD evaluation, "New Zealand has developed its own distinctive model of evaluation and assessment characterised by a high level of trust in schools and school professionals" (Nusche, Laveault, MacBeath, & Santiago, 2012, p. 3).

2.2.2.1   NCEA and teachers
Given the complexity of the NCEA system and the agency that teachers have within the system, a good understanding of all aspects of NCEA is essential to enable quality summative assessment practice. The requirements and effects of the high stakes internal school-based assessment for the NCEA on teachers has been documented and researched by a number of scholars (Brown et al., 2014; Hipkins, 2010, 2013; Hipkins & Neill, 2006; Hipkins & Vaughan, 2002; Shulruf, Hattie, &

Tumen, 2010; Wylie, 2013), including those with an interest in science education (Hipkins & Neill, 2006; Hume, 2006; Hume & Coll, 2009; Moeed 2013, 2015; Moeed & Hall, 2011). With respect to NCEA and teacher involvement, the literature focuses on the nature of NCEA courses, teachers' curriculum decision making and practice within NCEA courses, and their preparedness and confidence to assess for NCEA.

Secondary school NCEA courses are based on the NZC, but the course structure and the selection of standards used are not mandated so courses can be constructed by teachers with their local context in mind. This flexibility is a positive feature of NCEA, as courses can be purpose built for cohorts of students. However, the flexibility raises concerns regarding the potential narrowing of curriculum and compartmentalising of learning in NCEA courses, given that standards often assess discrete portions of the curriculum (Locke, 2005; Rawlins et al., 2005). School programmes or individual courses can become narrower or more fragmented when a subject is broken up into multiple ASs, and facets of a subject can be lost in NCEA courses (Brown et al., 2014). A study focusing on the influence of NCEA on the teaching and learning of mathematics, history, Japanese and French found that "washback" (i.e. the impact of assessment on teaching and learning) occurred for these subjects (Mizutani, Rubie-Davies, Hattie, & Philp, 2011). Although science was not a subject included in this study, it could be assumed that washback is a feature in NCEA science courses as well. For example, the Nature of Science (a requirement of the New Zealand science curriculum) may be lost in the fragmentation of science as the focus is narrowed to the ASs chosen for a course. Teachers need to be aware of these issues.

In a survey of the NCEA practices of a number of schools it was found that schools' programmes varied considerably both between schools and within schools (Hipkins et al., 2005). Hipkins and Neill (2006) found that schools did develop their own courses and that teachers removed standards from their courses where it suited them to do so. However, it has also been found that not all teachers are fully aware of the flexibility possible in NCEA course design (Hipkins, 2013). There are instances of some schools providing innovative programmes which cross traditional subject boundaries; for example, courses such as Agribusiness, Fitness for Living, Viticulture, Sea Sports and Pasifika, which use standards from

a variety of subject domains (Austin, 2015). Given the range of schools beginner teachers may end up working in, and the nature of NCEA courses, it is important that beginner teachers are aware of the risks and possibilities of NCEA as part of their developing assessment literacy for working with and within the NCEA system.

The assessment decisions teachers make in NCEA courses have been found to have both positive and negative outcomes for teaching and learning. In a study of science and mathematics teachers it was found that teachers reported an increased emphasis on levels and types of thinking, including critical thinking and metacognition, since NCEA was introduced (Hipkins & Neill, 2006). However, the same teachers reported a concern that fewer open-ended investigations were encouraged in NCEA courses. This potential problem was confirmed in research which found that science teachers did teach investigations in ways that were more about training students to complete an internal assessment than teaching them investigation processes (Hume, 2006; Hume & Coll, 2009, 2010; Moeed & Hall, 2011). In their case studies of assessment in Year 11 science, Hume and Coll (2009, 2010) found that NCEA practise tasks for science investigation were often very similar to the final summative task that contributed to the award of NCEA credits. This trend raises validity and reliability issues with respect to teachers' assessments and students' achievements in science investigation (Moeed & Hall, 2011).

Teachers are responsible for the assessment tasks they use and NZQA encourages the use of a wide range of assessment strategies and tasks in order to give students the best opportunities to provide evidence of their learning (NZQA, 2001). Naturally occurring evidence can be used for internal standards where appropriate, rather than having students complete stand-alone assessment tasks such as tests or assignments (Hipkins, 2012). The use of integrated assessment, where related or complementary outcomes from a range of standards can be assessed using the same task or evidence is also promoted by NZQA (NZQA, 2017c). One reason given for this is to reduce the overall time spent on assessment, and facilitate a more holistic approach to curriculum and assessment. Teachers therefore need to consider carefully the design of tasks they choose to summatively assess NCEA

standards, and they can use naturally occurring evidence or integrated assessment where it is appropriate.

For science teachers, up-to-date exemplar assessment materials are available online (http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/subjects/science), with other materials available from science teachers' associations and private resource developers (e.g., www.instant.org.nz). It is up to New Zealand classroom teachers to adapt these exemplar tasks so that they suit their own classes but still meet the requirements of the standards, which requires some knowledge of assessment design. In a study of teachers' responses to NCEA science investigation, Moeed and Hall (2011) found that science teachers were very reliant on simplified templates, indicating that teachers did not understand task design well. On the other hand, a number of innovative NCEA assessment practices have also been documented (Fastier, 2009; Jennings, 2008; Stoate, 2013). These include evidence gained through authentic tasks that allow students to demonstrate their learning in real-life situations.

As well as assessing students against the internal ASs and USs, secondary teachers prepare their students for externally examined ASs through the development and use of practise assessment tasks, usually in-school examinations. These practise summative assessment tasks provide standard-specific examination practise for students. The results of these practise tasks and examinations are used for reporting progress to parents and sometimes for formative use before the NZQA external examinations. Unhelpfully, these practise tasks are often referred to by teachers and students as 'formative assessments' or 'mock exams', irrespective of whether the results are used formatively by the teachers or students (Hume, 2006; Moeed & Hall, 2011).

The responsibility that New Zealand secondary teachers have for NCEA has led to the suggestion that this signals confidence in New Zealand teachers' abilities to assess (Starkey et al., 2009). However, individual teachers may or may not feel confident in the system and in their own abilities to assess within the system. Teacher stance towards the NCEA qualifications has been changing overall from teachers regarding it negatively to becoming more positive, although mathematics, science and computing teachers are over-represented in the negative

categories (Hipkins, 2013; Philips, 2003). Research into the effects of in-school moderation systems has found that the learning gained through the process of sharing materials and moderating students' work has led to teachers developing confidence in the use of standards, and to deeper collegial relationships (Barrett, 2008; Fastier, 2009; Hipkins, 2013). This confidence is not evident in beginning teachers, however (Anthony & Kane, 2008; Smith et al., 2013).

Research by Anthony and Kane (2008) has highlighted a perceived lack of preparedness by student teachers and beginner teachers for NCEA assessment. In their longitudinal mixed-method study of 855 secondary teacher graduates, over a third felt only 'somewhat prepared' or 'not at all well prepared' with respect to NCEA in general. The beginner teachers felt somewhat prepared in areas related to knowledge of assessment but many did not feel prepared in the use of assessment strategies, or in using the results of assessments in their subsequent planning. The majority of graduating teachers felt least prepared to develop and administer assessment protocols that reflect the requirements of NCEA. The beginner teachers believed that a lack of focus on NCEA curriculum achievement standards during initial teacher education (ITE) has contributed to their lack of understanding and confidence. The study went on to show that after their first and second years of teaching, most but not all teachers were developing confidence with respect to NCEA. Similarly, in a more recent study which followed a group of preservice teachers in their one year secondary ITE programme, it was found that the teachers lacked confidence and felt they had not had enough training in how to create reliable assessments, in the marking of assessments and in communicating assessment results (for both formative and summative assessment) (Smith et al., 2013). These findings would suggest it is important that NCEA assessment be a significant focus in preservice teacher education, as well in the support provided by schools for beginner teachers.

2.2.2.2   NCEA and students
Senior secondary students are able to choose their NCEA courses of study for each year of their senior secondary schooling. Most schools have some compulsory subjects (usually English and Mathematics), and many have prerequisites for entry into senior science courses, especially chemistry and physics courses, but notionally students have relative freedom to choose five or

six subjects. However, research has found that their choice is more limited than it needs to be. In a study involving mathematics and science teachers from 18 schools, Hipkins and Neill (2006) found that students deemed as 'capable' were offered courses with proportionally more externally assessed standards, whereas less able students were offered more internally assessed standards, seen to be more achievable given their flexible time frames and opportunities for re-assessment. Additionally, the Starpath Project at the University of Auckland found that schools with students from lower socio-economic groups were guided into NCEA subjects and standards that would potentially constrain their further academic options, such as university study. This could have the effect of perpetuating stereotypes for particular ethnic or socioeconomic subgroups (Jensen, McKinley, & Madyar, 2010; Madyar, McKinley, Jensen, & van der Merwe, 2009). Moreover, Shulruf, Hattie, and Tumen (2010) found that that a poor combination of subject choices hindered groups traditionally under-represented at university from gaining the credits necessary for university entrance. These studies highlight the need for students to be well informed and well advised so they choose the programme best suited for them. However, Meyer et al. (2009) found that students and their parents felt they needed more information, particularly from their teachers, about NCEA and the choices students had to make.

The New Zealand education system allows students to have input in curriculum planning within individual courses. However, this is not common practice. A survey of secondary schools in 2012 reported that only 13% of teachers surveyed reported students were asked for input in curriculum planning in their schools (Wylie, 2013). However, there is some evidence that students exercise agency by making their own choices with respect to individual NCEA assessments in strategic ways. For example, students can and do decide whether to complete or skip summative assessments for particular achievement standards, irrespective of teachers' guidance (Hipkins et al., 2005). This sort of action was of particular concern to teachers when the NCEA qualifications were first introduced. However, since that time teachers have responded in some cases by more readily accepting students' choices, and by reducing the overall number of standards per course in order to reduce students' and teachers' workloads (Fastier, 2009).

Students have been found to be motivated by the opportunity to gain higher than
Achieved grades, and the introduction of NCEA endorsements in 2007 and 2011
resulted in a positive response with more students aiming for a higher level of
achievement (Graham, Meyer, McKenzie, McClure, & Weir, 2010; Meyer, Weir
et al., 2007, 2009). However, it seems that less engaged students consider NCEA
as a credit-gathering exercise, doing just enough to gain the credits for each
standard (Graham et al., 2010; Meyer, McClure, Walkey, Weir, & McKenzie,
2009).

Research into views of Māori and Pacific students and parents has found that the
NCEA qualification "has the potential to support achievement by all students in a
manner culturally responsive to core values held by Pacific people and indigenous
Māori " (Graham et al., 2010, p. 152). This is because NCEA is not based on
competitive goals but can accommodate collective aspirations by encouraging
outcomes of collective excellence where individuals work together to "pull
everyone up to a higher level" (Graham et al., 2010, p. 149). However, it has been
found that these students are not taking full advantage of the flexibility and
opportunities available through the NCEA system. Overall, it is important that
New Zealand teachers understand the ways in which students can engage with
NCEA so they can support them in the most effective ways.

2.2.2.3   NCEA and quality assessment
Assurance of quality in summative assessment is important to all stakeholders,
giving them confidence that the qualifications based on the assessment are
trustworthy. As discussed above, two key moderation systems operate for NCEA:
a school-based social moderation system in which teachers moderate each other's
task design and marking, and an external NZQA facilitated moderation system in
which random selections of school assessment are checked by NZQA employees.
The benefits of social moderation for teachers' professional learning have been
established by a number of researchers (Hipkins, 2013; Smaill, 2013). However,
the workload demanded by the requirement of in-school moderation of NCEA has
been found to be challenging and a cause of concern for teachers, school leaders
and the unions representing them (Alison, 2005; Hipkins, 2013; Wylie, 2013).
The ongoing provision of exemplar materials consisting of real student work with
commentaries to help with moderation has been identified as being helpful for

teachers in developing their understanding of standards (Hipkins, 2012). School-based critique of assessment materials and moderation is seen as the most important quality check within the NCEA system.

External moderation is carried out by NZQA-employed moderators as a further check for quality. Feedback from external moderation must be incorporated into the schools' assessment systems according to the *Assessment (including Examination) Rules for Schools with Consent to Assess 2015* (NZQA, 2015b). Additionally, NZQA audits schools with a Managing National Assessment (MNA) visit regularly to ensure they are meeting all requirements. In this audit the management of the schools' assessment practice, moderation, data and communication is scrutinised, and an MNA report is produced that summarises the findings and details any actions required for the school to maintain its consent to assess.

A recognition of the need for quality in NCEA summative assessment, including understanding NZQA policy, practices and procedures to ensure quality such as the use of moderation systems, requires a high level of assessment literacy from all teachers, including beginner teachers in New Zealand schools.

2.2.2.4 Summary
The purposes of summative assessment, its quality, forms and impacts have been reviewed, particularly with respect to the summative assessment used in New Zealand secondary schools. Beginner teachers' development includes building their knowledge of the range of uses of summative assessment, and developing an awareness of the considerations they need to have when choosing when, where and how to assess students. This is of particular importance for high stakes assessment such as NCEA in New Zealand.

2.3 Assessment literacy

This section considers teachers' understanding and use of assessment. Numerous terms are used in the literature when describing what teachers need to be able to assess learners and learning within any specific context. For the purposes of this thesis, the term *assessment literacy* is used to describe the phenomenon under investigation. This section reviews definitions of assessment literacy and other related and somewhat overlapping terms and concepts found in the research

literature (Section 2.3.1) and explains how assessment literacy is defined for this study (Section 2.3.2).

2.3.1 Defining assessment literacy

The international movement towards accountability in education together with the research evidence that confirms the benefits of assessment-informed teaching mean that assessment literacy is now considered a fundamental component of what teachers need (Popham, 2009; Xu & Brown, 2016). As Popham (2009) comments, "educators' inadequate knowledge … can cripple the quality of education. Assessment literacy is seen as a sine qua non for today's competent educator" (p. 4). The term assessment literacy is used by a wide range of researchers and authors to describe what teachers need to know about assessment matters (Abell & Siegel, 2011; Klenowski & Wyatt-Smith, 2014; Lomax, 1996; Lukin, Bandalos, Eckhout, & Mickelson, 2004; Mertler & Campbell, 2005; Popham, 2009; Siegel & Wissehr, 2011; Stiggins, 1995; Xu & Brown, 2016).

Assessment literacy has been defined as an understanding of the principles of sound assessment (Stiggins, 1995, 2002). Stiggins (1995) commented that "[a]ssessment literates know the difference between sound and unsound assessment. They are not intimidated by the sometimes mysterious and always daunting technical world of assessment" (p. 240). Stiggins goes on to explain that assessment literate educators

> come to any assessment knowing what they are assessing, why they are doing so, how best to assess the achievement of interest, how to generate examples of performance, what can go wrong, and how to prevent these problems before they occur. Most important, those who are truly sensitive to the potential negative influences of inaccurate assessment never permit students to be put in a situation where their achievement might be mismeasured. (Stiggins, 1995, p. 240)

Adding to this, Popham's (2011) definition that "assessment literacy consists of an individual's *understandings of the fundamental assessment concepts and procedures* deemed *likely to influence educational decisions*" (p. 267, emphasis in original) draws our attention to knowledge and decision making aspects. Likewise, DeLuca and Klinger (2010) draw attention to both theoretical and practical knowledge as well as an underpinning framework when they explain assessment literacy as involving "the understanding and appropriate use of

assessment practices along with the knowledge of the theoretical and philosophical underpinnings in the measurement of students' learning" (p. 419).

The New Zealand Ministry of Education appears to have drawn from Stiggins' work as well as that of Absolum et al. (2009) to describe assessment literacy as follows:

> Assessment literacy is the possession of knowledge about the basic principles of sound assessment practice, including its terminology, the development and use of assessment methodologies and techniques, and familiarity with standards of quality in assessment. Teacher and students understand that assessment, both informal and formal, is central to teaching and learning and use it formatively at all times to determine where individuals and groups are with their learning. Teachers have the capability to gather dependable information, aggregate and analyse it, and use it to further teaching and learning for individuals and groups of students. (Te Kete Ipurangi, n.d., para. 1)

Thus we can see descriptions of assessment literacy are similar, mostly couched in terms of the knowledge of aspects of assessment theory and practice and of the potential consequences of assessment, with the implication being that this knowledge allows teachers to discern between sound and unsound assessment and articulate reasons for making specific decisions and actions.

More recently the concept of assessment literacy has been theorised to include the importance and effects of teachers' views of learning, and their assessment values and principles (Abell & Siegel, 2011; Brookhart, 2011; Gottheiner & Siegel, 2012). This view encompasses beliefs, knowledge and skills, and acknowledges both theory (theoretical principles for assessment) and practice (practical methods of use in a classroom), which evolve at the same time and from each other (Kahl, Hofman, & Bryant, 2013; Leighton, Gokiert, Cor, & Heffernan, 2010; Mertler & Campbell, 2005; Siegel & Wissehr, 2011; Xu & Brown, 2016).

From a sociocultural perspective, assessment literacy is viewed as a dynamic social practice in which contextual and cultural influences interact and are acknowledged (Adie, 2011; Gee, 2008; Gipps, 1999; Pryor & Crossouard, 2008; Willis et al., 2013). The following definition has been offered by Willis et al. (2013):

> Assessment literacy is a dynamic context dependent social practice that involves teachers articulating and negotiating classroom and cultural knowledges with one another and with learners, in the initiation, development and practice of assessment to achieve the learning goals of students. (p. 242)

This definition shifts the focus on assessment literacy from a fixed set of capabilities held by teachers or individuals, to the concept of assessment literacies. Given the complex contexts involved in classroom assessment and the multiple discourses in operation, this definition captures the fluidity of the processes involved during assessment (Hay & Penney, 2012; Willis et al., 2013).

Klenowski and Wyatt-Smith (2014) explain that for teachers to be assessment literate they

> need to have a repertoire of skills and understandings to design quality assessments and to use achievement standards and evidence as a means by which to discern, monitor and improve learning as well as judge the qualities of student work" (p. 2).

These authors identify eight essential professional capabilities and attributes that contribute to teachers' assessment literacies. In brief, it is argued that teachers need well developed curriculum knowledge; to be able to self-identify and review assumptions to verify their suitability; to be able to align assessment evidence collection with learning objectives; to consider the range and balance of tasks set for students; to be aware and responsive to student diversity; to develop skills in communicating cognitive expectations and purposes of set tasks; skills in communicating the characteristics of quality in set tasks; and a willingness to reflect on and share their judgements with other teachers. This repertoire of knowledge, capabilities and dispositions is called upon when teachers are front-ending assessment, that is when they are "directly connecting the intended curriculum with decisions about the information to be collected in order to know what learning is occurring in the classroom and how best to tailor teaching and learning opportunities to enable success for all learners" (Klenowski & Wyatt-Smith, 2014, p. 55). It is argued that the knowledge, capabilities and dispositions identified will support valid assessment.

Assessment competence is a term with a similar meaning to assessment literacy used by Black et al. (2010) in their investigation of teachers' summative

assessment practices. This term is also used by researchers working in the area of language teaching and has resulted in the development of descriptors for six levels of diagnostic competence (Edelenbos & Kubanek-German, 2004). Xu and Liu (2009), also working in the area of language learning, use the term *assessment knowledge*. These researchers use the concept of teachers' professional knowledge landscapes (Connelly & Clandinin, 1995) as a framework within which to place teacher assessment knowledge, and acknowledge temporality, sociality and place and their effects on teachers as they build assessment knowledge.

The term *assessment expertise* is used by some authors in an attempt to capture teacher growth towards more sophisticated assessment practices (Gearhart et al., 2006; Lyon, 2013a, 2013b, 2013c). Assessment expertise is defined as "the relationship between assessment understanding and assessment facility" (Gearhart et al., 2006, p. 259) so it focuses not only on the knowledge of assessment but on its enactment. Understanding assessment concepts refers to teachers' knowledge of what constitutes quality assessment, such as quality tools for gathering evidence and sound interpretations of the evidence, as well as an understanding of the requirement that all components of assessment must be aligned and embedded in a system to provide coordinated information and action. Assessment facility refers to how they use this understanding in the classroom — their planning, using and reflecting on assessment tasks and the evidence of learning these produce.

Assessment capability is a term recently adopted in New Zealand where it has been defined as being "able and motivated to access, interpret and use information from quality assessment in ways that affirm or further learning" (Absolum et al., 2009, p. 6). In their forward-thinking policy advice paper to the New Zealand Ministry of Education, *Directions for Assessment in New Zealand*, Absolum et al. (2009) position the development of student assessment capability as well as teacher assessment capability as a central premise, and state that students will only develop assessment capability if their teachers and school leaders have both genuine assessment capability and appropriate resourcing. The terms 'able and motivated' within the definition of assessment capability signal that both the knowledge and the disposition of teachers are integral to their use of assessment. Teachers need a deep knowledge of the curriculum, pedagogy and learning, and also require motivation to develop learning partnerships with students;

partnerships which acknowledge the effects of assessment activities on students. Assessment capable teachers are those who

> understand how students can use and value assessment as a powerful means of furthering their own learning. As the experts in the learning partnership, teachers need to take the lead in all assessment that students cannot manage without support. But they need to do so in ways that encourage students to feel deeply accountable for their own progress and support them to become motivated, effective, self regulating learners. (Absolum et al., 2009, p. 24)

Thus there are a range of terminologies used in literature to describe what teachers need to assess, and how teachers enact this in assessment contexts. This literature defines and discusses assessment literacy in general, rather than summative assessment literacy specifically. There are considerable overlaps in the different definitions, and this means that often very similar ideas are described using different terminology. These ideas include teachers having assessment knowledge and an understanding of assessment processes, as well as an understanding of the underpinning principles of quality assessment and the ability to apply these in assessment decision making. The term *assessment literacy* has been chosen for this study because of its widespread use internationally to describe my area of interest. Definitions of assessment literacy encompass theoretical and practical aspects of assessment, whilst acknowledging that assessment is an active process that occurs within contexts which interact with the people and processes involved.

2.3.2 Assessment literacy in this study

For the purposes of this study, the definition of assessment literacy takes account of both constructivist and sociocultural perspectives. These have played an important role in research that focuses on science and mathematics education and teacher development over the last three decades as researchers have recognised that a combination of these approaches can be helpful in explaining and discussing phenomena (Buxton et al., 2015; Chazan, Brantlinger, Clark, & Edwards, 2013; Cobb & Bauersfeld, 1995; Cobb, Jaworski, & Presmeg, 1996; Cobb & Yackel, 1996; Hodge & Cobb, 2016; Johnson, Nyamekye, Chazan, & Rosenthal, 2013; McCaslin & Legg Burross, 2011; McDonald, Warren, & DeVries, 2011; Tabor, 2008). The proposition is that using lenses that attend to the individual dimensions of experience and to the social and cultural dimensions

enables a richer understanding of teaching and learning (Cobb et al., 1996; Kieran, 2000); in this view the relationship between cognitive and sociocultural perspectives can be seen as complementary and interwoven.

In this study the two following definitions are therefore adopted. Stiggins' (1995) definition that assessment literate educators

> come to any assessment knowing what they are assessing, why they are doing so, how best to assess the achievement of interest, how to generate examples of performance, what can go wrong, and how to prevent these problems before they occur. Most important, those who are truly sensitive to the potential negative influences of inaccurate assessment never permit students to be put in a situation where their achievement might be mismeasured (Stiggins, 1995, p. 240)

focuses on the individual's understanding of the principles of sound assessment, and is used to account for the cognitive perspective. Stiggins' individual cognitivist view allows for the development of assessment literacy to be tracked for individual teachers. Secondly, the definition by Willis et al. (2013) that

> assessment literacy is a dynamic context dependent social practice that involves teachers articulating and negotiating classroom and cultural knowledges with one another and with learners, in the initiation, development and practice of assessment to achieve the learning goals of students (p. 242)

has been developed from a sociocultural perspective, and is used to enable an understanding of the contextual elements involved during teachers' development of assessment literacy. Therefore for the purposes of this study, summative assessment literate teachers will be viewed as those individuals who understand the principles of sound assessment and who use summative assessment knowledge and practices appropriately. They will also be viewed as members of a community in which shared knowledges and particular practices of summative assessment have been developed and are valued and used for a variety of purposes in ways that are dependable and seen as fair and context-appropriate. By viewing teachers' assessment literacy through individual and sociocultural lenses, a richer description of their development of summative assessment literacy is hoped to ensue.

2.4 Summary of assessment and assessment literacy

In this chapter literature on assessment has been reviewed in order to provide a basis for the study. In particular, studies which focus on aspects of summative assessment both internationally and within New Zealand have been carefully considered. A thorough understanding of the principles, purposes, quality and impact of summative assessment is necessary for New Zealand secondary science teachers, given their involvement in summative assessment with their students, both for reporting and in-school use, and for high stakes qualifications. New Zealand secondary science teachers are seen as part of the wider professional community to which they contribute and develop shared understandings and practice. This knowledge and involvement has been described as their summative assessment literacy.

The next chapter, Chapter 3, reviews literature that focuses on teacher professional knowledge and assessment literacy development. It reviews research on teacher development over time and the influences that affect developing teachers. This chapter then presents models of teacher knowledge with a focus on knowledge for assessment. Summative assessment literacy is positioned as an essential component of teachers' professional knowledge. Following this, a section on teacher conceptions of assessment provides background for the study. Research regarding the development of teachers' summative assessment is then presented followed by a focus on New Zealand specific literature.

Chapter 3: Teacher professional knowledge and assessment literacy development

In this chapter literature that focuses on teacher development is reviewed (Section 3.1). This is followed by a review of models of teacher knowledge with a focus on knowledge for assessment (Section 3.2). Next, the chapter focuses on the development of teacher assessment literacy in general terms and in the New Zealand context (Sections 3.3 and 3.4). Finally, the chapter is summarised (Section 3.5). Section 3.6 looks across Chapters 2 and 3 to give context for the research questions, and states the questions.

3.1 Developing as a teacher

Becoming a teacher is a process that involves both personal and professional growth. This section considers cognitive and sociocultural perspectives in teacher development (Section 3.1.1), the place and influence of ITE in teacher development (Section 3.1.2), and other non-ITE factors influencing teacher development (Section 3.1.3). It then moves on to review studies regarding teacher development over time (Section 3.1.4).

3.1.1 Cognitive and sociocultural perspectives in teacher development

A number of important factors influence people as they develop as teachers. Some of these factors are brought by the individuals themselves, stemming from their life histories, their personal conceptions and values, abilities, and willingness to reflect on their practice. Personal factors such as these affect what teachers notice and prioritise, and their ability to learn and to process their experiences and construct a knowledge base for teaching and educational decision making (Campbell, 2008; Eraut, 1994; Gess-Newsome, 2015; Kelly, 2006). Other influences on teacher development are external to the teacher, in that they arise from formal aspects of teacher education, including input from university courses, from practicum experiences, from interactions with other practitioners and students, and from the contexts (political, historical, physical) in the communities in which teachers work (Gess-Newsome, 2015; Hill, Grudnoff, & Ell, 2011; Loughran, 2006; Mansour, EL-Deghaidy, Alshamrani, & Aldahmash, 2014; Wilson & Demetriou, 2007). In order to capture this diversity of influences,

teaching has been theorised from a cognitive and a sociocultural perspective, although a sociocultural perspective is dominant currently (Anderson, 2007; Bell, 2011; Borko, 2004; Postlethwaite & Haggarty, 2012).

The sociocultural perspective advocates that knowledge is distributed across people and settings (Lave, 1996; Lave & Wenger, 1991). It acknowledges that teachers bring their own perspectives to the social context in which they work (Woods & Jeffrey, 2002). Hence, although learning can be seen as an individual experience, from this view point it is understood to be deeply rooted in the community of practice in which knowledge and values is distributed (Semin & Smith, 2013). Teacher learning is viewed as the movement from peripheral or novice participation in the specific environment of the school and classroom, to full or expert participation, along with an increased alignment with the ways of knowing and thinking held common in the community of practice of teachers in a particular school. As teachers develop expertise or capability and they become more knowledgeable about the practices of teaching, or in the case of this study summative assessment literacy, their zone of proximal development (Vygotsky, 1978) extends outwards to encompass new ideas. The scaffolding or support teachers receive from others enables them to build their expertise and become confident in its use. In this way sociocultural perspectives take account of the social interactions teachers have within and across contexts, where the interplay of these interactions is thought to shape their learning (Postlethwaite & Haggarty, 2012). In summarising the sociocultural perspective in relation to teacher development, Mansour et al. (2014) argue that "teachers' learning, their practice, their development are seen as dynamic, multi-layered processes constituted in dialectic relationship between micro individual and interactional elements of teachers' learning and larger social, political and economic macro contexts of practice" (p. 952).

A number of researchers have described how teachers learn both personally as individuals and within communities and contexts in ways that acknowledge both cognitive and sociocultural aspects of learning and change. Examples of such framing will be presented next.

Shulman and Shulman (2004) offered a new frame as a response to different reactions from teachers who experienced the same professional learning programme. This provides a way of reconciling ideas of PCK, which was seen as strictly cognitive and individual, with the disciplined critical and collective reflection and case study work through which the concern for teacher learning moved from a focus on the individual to a conception of learning within the broader context of community. The Shulman's argue for "the importance of a communal cluster that includes deliberation, collaboration, reciprocal scaffolding, and distributing expertise" (Shulman & Shulman, 2004, p. 265), thus acknowledging that teacher learning is more than an individual endeavour.

The Shulman and Shulman (2004) perspective aligns with the findings of earlier work in the Learning in Science Project (Teacher Development), which was a three-year study on teacher development in New Zealand science teachers (Bell and Gilbert, 1994). From their work, Bell and Gilbert (1994) identified the three main types of development for teachers involved in their research as social, personal and professional development. They proposed that these three foci acknowledged both the cognitive and the social aspects of learning and change. As Bell (2011) suggested more recently:

> The sociocultural perspective is important as it asserts that, to understand teaching, we need to consider the teacher's mind (cognition, affect and conation), the teacher's action and the context in which the teaching occurs. To neglect one of these (mind, action and context) is to only partially understand teaching. (p. xi)

Zembylas (2007) has similarly argued for the need to consider teacher affect in teacher development. He suggests the need for a broadening of the conception of PCK to allow for consideration of emotion, intuitive and tacit dimensions of teacher knowledge in both individual and collective teacher development. Zembylas defined teacher emotional knowledge as:

> a teacher's knowledge about/from his or her emotional experiences with respect to one's self, others (e.g. students, colleagues), and the wider social and political context in which teaching and learning takes place. (p. 355)

He used case studies to illustrate how teacher emotional knowledge could be seen as an inextricable component of the amalgam of PCK, operating in three planes: individual, relational, and socio-political.

The idea that teacher learning simultaneously occurs at an individual level and is situated in and shaped by the context, embedded as it is in the community of practice within a school is also supported by Wilson and Demetriou (2007). According to these researchers learning is shaped by the reciprocity between the learning context in a particular school setting and an individual teacher's disposition to learn and interest in learning about practice (see also Billett, Fenwick, & Somerville, 2006; Wallace & Loughran, 2012; Wilson & Demetriou, 2007). With regard to becoming a science teacher, Luft, Dubois, Nixon, and Campbell's (2015) overarching conclusions highlighted both the individual and social components of teacher development and the need for ITE programmes that align with well school-based induction programmes.

The literature reviewed in this section provides strong theoretical perspectives from which to consider teacher development. In this study, a combination perspectives is employed, so that both individual knowledge and PCK, as well as social components of teacher development can be used as complementary lenses.

ITE programmes play a central role in preparing teachers for their work. Literature regarding the place and influence of ITE as a factor contributing to "becoming a teacher" is reviewed in the following section.

3.1.2 Place and influence of ITE in teacher development

The kind of ITE programme that preservice teachers encounter will determine their learning experience (Adoniou, 2013; Darling-Hammond, 2006a; Kennedy, 1999). In the 1970s and 1980s it was asserted that ITE had a limited impact on classroom practice (Lortie, 1975; Veenman, 1984), with some researchers reporting a washout effect once teachers started working i.e., the effects of ITE disappeared over time. However, ongoing research shows that ITE programmes do make a difference to the teaching skills of teachers by providing a knowledge base from which to develop pedagogy (Berliner, 2001; Darling-Hammond, 2000; Hoffman et al., 2005; Rice, 2003). It has been found that over time teachers move from being a novice in aspects of knowledge about teaching and learning to

developing a more informed and experienced approach to their work (Darling-Hammond, 2000; Edwards, 2009; Gearhart & Osmundson, 2009; Graham, 2005; Hill et al., 2011; Levin & He, 2008).

Much of the literature that discusses ITE focuses on its two major components: university-based course work (usually seen as theory work), and school-based practicum experiences (usually seen as practical work) (Adoniou, 2013). Darling-Hammond (2006b) argues that three critical components which bring these together create stronger and more effective teacher education programmes:

> tight coherence and integration among courses and between course work and clinical work in schools, extensive and intensely supervised clinical work integrated with course work using pedagogies that link theory and practice, and closer, proactive relationships with schools that serve diverse learners effectively and develop and model good teaching. (p. 300)

University course work serves a number of purposes within ITE. There is a recognised body of knowledge as well a range of skills and dispositions that university courses focus on with beginner teachers. Through course work they are exposed to ideas that relate to learning theories, pedagogy, assessment, curriculum, professionalism as well as more philosophical and theoretical aspects of education that link to subjects such as philosophy, history, sociology, anthropology and politics (Rogers, 2011; Stuart & Thurlow, 2000). Involvement in such courses can help preservice teachers to develop their pedagogical reasoning. Youngs and Bird's (2010) investigation into the development of pedagogical reasoning among secondary teaching candidates, for example, demonstrated the importance of the nature of assessment tasks embedded in their ITE programme. This study provided evidence that the use of high-gain course-embedded projects was one way to move teachers towards mastery in elements of pedagogy. Similarly the use of CoRes and PaP-eRs (Loughran, Berry, & Mulhall, 2006) in ITE courses has been shown to help novice science teachers understand what PCK in science might involve (Hume, 2010; Hume & Berry, 2010). Another study investigating specific ways in which to help preservice teachers learn about assessment processes by involving them in social moderation for their peers, found that although the process was viewed positively by the teachers, some still struggled to understand standards referenced assessment (Grainger & Adie, 2014).

This study highlighted the need for ongoing, scaffolded preparation for preservice teachers before they engage in peer assessment, grading and moderation.

Practical teaching experiences are another essential element of ITE. Practicum gives student teachers knowledge *of* teaching as opposed to knowledge *about* teaching and is the context in which teachers develop personal practical theory and a teacher identity (Smith & Lev-Ari, 2005). The effectiveness of the learning that occurs on practicum is arguably dependent on the integration of theory and practical contexts, which in turn requires careful planning for links between coursework and practicum (Adoniou, 2013; Allen, 2009; Bullough, 1992; Cheng, Tang, & Cheng, 2012; Darling-Hammond, 2006b). In a detailed case study presented to report on affordances and constraints provided by context in teacher education, Smith and Jang (2011) found that a practicum context provided affordances to enable a teacher participant to move towards more central participation (Lave & Wenger, 1991). In contrast, others have found that there are times when practicum provides constraints or tensions for preservice teachers as they try to reconcile what they are told is good practice in their university courses with what they experience in schools. The suggestion is that this sort of experience can result in student teachers being socialised to the status quo (Pugach, 1992; Taber, et al., 2011; Winterbottom, Brindley et al., 2008; Winterbottom, Taber et al., 2008).

The associate teachers (ATs) or mentoring teachers in schools who work with beginner teachers are very important contributors to beginner teachers learning to teach. These teachers can affect beginner teachers' sense of adequacy and confidence, their abilities to reflect on theory and practice and the conceptions they hold of their identity as a teacher (Bullough, 1992; Graham, 2005; Grudnoff, 2007; Langdon, 2007). In a review of research on ITE in New Zealand, Cameron and Baker (2004) identified the important role that ATs play for teachers on practicum. They commented on the need for a closer match between the ITE programme outcomes and expectations of ATs, as the sometimes poor supervision and feedback provided to preservice teachers, and the pressures that preservice teachers felt, were seen to reduce the chances of effective teacher development. More generally, it has been found that good communication and a collaborative approach to practicum outcomes and processes mean preservice teachers are more

likely to gain from the experiences, but that there are times when learning is compromised during practica (Gibbs, 1995; Haigh, Pinder, & McDonald, 2008; Murray, 2013; Ovens, 1996). For example, Haigh et al. (2008) found that the relationships teachers had with ATs, visiting lecturers and students could enhance or inhibit their learning to teach.

Rogers (2011) argues for a coherent view of teachers' ITE and the learning experiences they encounter in their programme whilst acknowledging that teacher-educators must also introduce alternative views and challenges because

> the vitality of teacher education rests with its capacity as an agency of change, not in recycling trusted and familiar routines; and at the heart of a change-mechanism is learning, and with that, an expression of personal values and beliefs about teaching. (Rogers, 2011, p. 250)

Like Loughran (2006), Rogers (2011) advocates the role of teacher educators to admit complexities, encourage critical thought and move preservice teachers away from the comfortable, familiar understandings of learning and teaching and the idea that they are learning the 'right way' to teach. This approach to ITE of challenging student teachers' beliefs, particularly unexamined beliefs, enables preservice teachers to examine and theorise their teaching, as well as develop their knowledge base (Stuart & Thurlow, 2000). Beginner teachers' epistemological growth has been identified as a key component in their development as teachers and on their teacher identity. Studies have shown that this growth is facilitated by a specialised disciplinary component of their ITE programme (Harteis, Gruber, & Lehner, 2006; Rogers, 2011). In the study reported in this thesis, university course work including a specialised science disciplinary component is a large component of the ITE programme. Beginner teachers' university experiences purport to contribute to teachers' assessment literacy (Deluca &Klinger, 2010), and so these are investigated.

Beginner teachers have been found to claim that their ITE programmes were not practical enough and that their practicum experiences were where most learning occurred (Sjolie, 2014; Smith & Lev-Ari, 2005). However, the common perception that ITE is too theoretical may be more about the preservice teachers' limited conception of the nature of theory and their ambivalent relationship with it (Darling-Hammond, 2010; Sjolie, 2014). For example, Allen (2009) reported on

the views of beginner teachers who had been involved in an ITE programme which set out to be purposeful in the way the programme addressed the theory-practice gap. Even though the beginner teachers valued both the theoretical course work and the practicum aspects of their ITE whilst studying, they privileged the learning they did on practicum, and associated good practice with more experienced teachers in the workplace. On the other hand, when Sjolie (2014) investigated preservice teachers' views of the role of theory in teacher education, she found they held a dichotomous view of theory and practice and their conceptions of theory were narrow, but they held nuanced views on the purposes for theory, moving beyond the expected 'theory for practice' view. For this study, it is assumed beginner teachers will be developing summative assessment literacy as they participate in both university courses and practicum experiences so it is important to observe the contributions each make, and the teachers' conceptions of how this happens.

Graham (2005) described an ITE programme that integrated theory and practice by working with preservice teachers and their school based mentors, using an assessment-based approach in which mentors were required to model backward planning to show preservice teachers how formative assessment data was used to inform their pedagogical decisions. It appeared that this integrated approach allowed preservice teachers to focus more clearly on their assessment practice as they could see how theories discussed on campus were understood and enacted in the context of the classroom. The importance of professional dialogue was emphasised in this study. Cheng, Tang, and Cheng (2012) highlighted the very important role of teacher educators in supporting preservice teachers to move towards a reflective theorising approach that practicalised theoretical knowledge and drew together the elements of their ITE programme.

Postlethwaite and Haggarty (2012), in their investigation of the way science and mathematics preservice teachers think about learning to teach, suggest that teachers use "progressive filtering" (p. 278) as a way of choosing what to focus on over the course of their teacher education. They argue the two significant filters are the teachers' history and their school experience (as preservice teachers). This could explain why "university programs continue to graduate teachers who are overconfident and under competent when it comes to

summarizing achievement and using assessment information to promote improved student learning" (Moss, 2013, p. 252). The development of summative assessment literacy sits within overall beginner teacher development so it is important to consider the impacts of ITE on development.

3.1.3   Other factors influencing teacher development

There are a number of factors besides those stemming from ITE that affect beginner teachers' learning and development. Teachers learn within a wider sociocultural context, their previous experiences and life histories shape their views and their learning, and factors from the affective domain, including emotion, have also been shown to influence teacher learning and decision making.

3.1.3.1   *The educational context*

The effects of the sociocultural contexts that teachers find themselves in have been explored in order to consider affordances and constraints that these contexts provide for teachers. Detailed case studies are often provided to illustrate the effects of contextual policy priorities, culture, and the classroom context (Brown, 2011; Brown, Lake, & Matters, 2011; Mahuika et al., 2011; McGinnis, Parker, & Graeber, 2004; Rogoff, 2008; Shulman & Shulman, 2004). For example, in the USA, Bullough (1992) described teacher development through the examination of beginning English teachers, and found that factors such as school policy, mentor teachers and timetabling could help or severely slow or limit teachers' development. Similarly, McGinnis et al. (2004) found that beginning mathematics and science teachers reacted in different ways depending on their perceived school culture, and thus beginning teachers flourished where they found support, or resisted, moved or quit teaching if less supportive school cultures were experienced. In another US case study, Smith and Jang (2011) found that a beginner teacher's developmental paths were affected by the contexts in which learning occurred. By situating the teacher's own thinking in the context of their lives, it is possible to understand the teacher's progress and complications in a more nuanced way.

As an example of the New Zealand educational context, New Zealand's assessment policy focus on formative assessment has been found to affect teacher conceptions of assessment. When comparing his studies with those from other scholars, Brown (2011) commented that "New Zealand teacher conceptions of

75

assessment reflect the much greater emphasis in New Zealand's educational culture for assessing students to improve learning, even in the secondary school qualifications system" (p. 63). Brown theorised that "teachers' conceptions of assessment are ecologically rational in that they reflect the legal, cultural or social priorities placed on assessment for their work environment" (Brown, 2011, p. 65).

3.1.3.2   Teachers' conceptions and background experiences

The term conception is used in this section "as a more general mental structure, encompassing beliefs, meanings, concepts, propositions, rules, mental images, preferences, and the like" (Thompson, 1992, p. 130). Teachers' conceptions about teaching and learning include their beliefs about teaching and learning, and are largely derived from experiences over time, and start to form early. They have been found to act as a filter for further knowledge and ideas about teaching and learning and to persist even when the holder is presented with compelling new learning (Bransford, Brown, & Cocking, 2000; Dixon & Haigh, 2009; Dixon, Hawe, & Parr, 2011; Edwards, 2003; Lortie, 1975; Postlethwaite & Haggarty, 2012; Taber et al., 2011). For example, the beliefs that beginner teachers hold when they enter ITE have been shown to substantially affect their knowledge acquisition and interpretation, task definition and selection, interpretation of course content, comprehension monitoring, and subsequent judgements and actions (Chant, Heafner, & Bennett, 2004; Fang, 1996; Pajares, 1992; Renzaglia, Hutchins, & Lee, 1997; Richardson, 2003). Beginner teachers' assessment practices are affected by their conceptions of how students develop and demonstrate skills and knowledge in a subject domain, and how and why teachers should assess student knowledge, performance and skills (Brown & Michaelides, 2011; Fang, 1996; Harris & Brown, 2009; Magnusson, Krajcik, & Borko, 1999; Mislevy, 1993; Pellegrino, Chudowsky, & Glaser, 2001). As Corrigan et al. (2013) point out, teachers' personal conceptions of assessment and "knowledge of worth" (p. 3) contribute to their approach to classroom assessment. It is theorised that beliefs formed earlier in life are more connected than subsequent beliefs and are therefore very hard but not impossible to change in preservice teachers, so it is important that naïve or tacit beliefs are made explicit and challenged within ITE programmes (Horwitz, 1985; Levin & He, 2008; Loughran, 2006; Pajares, 1992; Stuart & Thurlow, 2000).

The influence of their conceptions does not stop once teachers complete their ITE, but instead the influence continues throughout their professional lives (Cronin-Jones, 1991; Gess-Newsome, 2015; Kagan, 1992; Smith, 1996). However, the school context and culture in which the teacher works has been found to be very influential in the ongoing development of teacher beliefs. In her recent summary of literature on teacher beliefs, Levin (2014) argued that

> in fact, attending to context is paramount to understanding the development of beliefs because teachers' beliefs and actions cannot be separated from situations in which they occur; including the larger social, political, and economic climate, as well as the immediate school context. (p. 51)

Some school contexts allow for alignment between teachers' beliefs and actions but others do not. For example, Hill (2000) found that conflict resulted for teachers in their use of assessment as they sought to meet the competing demands of the school. Similarly, James and Pedder (2006) and Hipkins and Neill (2013) found that as a result of external constraints derived from educational policy, teachers' beliefs and practice did not always align.

Many studies have investigated the effect of teacher conceptions or beliefs on their assessment practice. In a study comparing the conceptions held by primary and secondary school teachers in New Zealand, Brown (2011) found that both primary and secondary teachers responded in very similar ways. Brown suggested this could be a consequence of both kinds of schools working in a low stakes national policy framework given that both primary and secondary teachers reported strong conceptions of assessment for improvement. Secondary teachers were found to agree more with using assessment to evaluate or grade students. Further research focused on the conceptions of teachers involved in NCEA assessment. Yates and Johnston (2017) found that the conceptions of secondary teachers were more closely aligned to teachers in high-stakes qualifications jurisdictions. *Assessment for qualifications* and *improvement* conceptions were interpreted in the context of school based assessment for NCEA.

Beliefs about learning or learning theory have also been linked to teacher assessment practices. Reviews such as those by Shepard (2000) and Earl (2013) have provided a historical overview to illustrate the change in theories of learning

over time, and how this links to changes in assessment. The link between teachers' views of learning and their conceptions of assessment has been investigated by a number of researchers, and although one might expect a direct correlation between these, this has not always been evident in the international literature. For example, in Taiwan a correlation between teachers' conceptions of learning science and their conceptions of how it should be assessed was found (Lin, Lee, & Tsai, 2014). In contrast, Wang, Kao, and Lin (2010) found that Taiwanese preservice elementary teachers' views of learning science and their assessment practice were in conflict because of the cultural influences at play in a society where success in national examinations is highly valued. In her work in Spain, Remesal (2011) highlighted the complex nature of assessment, which meant that simple relationships were not decipherable, and instead the teachers' conceptions were found to sit within and be influenced by a broader societal context. This was also found to be the case in a study in the UK where conceptions of assessment were also linked to subject speciality areas (Winterbottom, Taber et al., 2008).

Closely linked to beliefs, background and prior experiences have been shown to have a marked influence on teacher learning and teacher identity (Aitken & Mildon, 1992; Daly, 2008; Maclellan, 2004; Xu & Liu, 2009). The use of narrative inquiry in particular has enabled examples of the complex effects of these experiences to be examined, as teachers grapple with what they are learning and doing and compare this to their earlier personal experiences. For example, Xu and Liu (2009) found that a teacher's prior experience of being assessed had ongoing effects on her knowledge construction with respect to assessment and her assessment practice. The authors in this study emphasised the importance of prior experiences and suggested that interrogation of prior assessment experiences is needed to allow teachers to reconstruct their personal practical knowledge. Similarly, Edwards and Edwards (2016) document the effects of a New Zealand teacher's assessment experiences as a school student on his current assessment preferences as a beginner teacher. The study showed the lasting impact of cultural and linguistic experiences on the teacher in his early years as a student in Māori medium schooling. He attributed his current understanding of best-practice assessment to the priorities afforded in this early school environment as a student.

Even though key principles underpinning quality assessment have been identified and documented by researchers and policy makers, teachers conceptualise quality assessment in different ways, leading to their own practices. The science education literature indicates that the development of an integrated understanding of teaching and assessment is idiosyncratic for science teachers, which might further explain the variety of conceptions of quality assessment teachers bring to science classrooms (Aydin, Demirdogen, Akin, Uzuntirkaki-Kondakci, & Tarkin, 2015; Brown & Michaelides, 2011; Brown, Lake, & Matters, 2011; Grossman, 1990; Harris & Brown, 2009; Park & Chen, 2012; Park & Oliver, 2007). It has been suggested that the idiosyncratic nature of conception development may be as a result of the individual and personal nature of PCK, as well as teachers' orientations, conceptualisations, characteristics, and the complex subject matter knowledge (Aydin et al., 2015; Gess-Newsome, 2015; Lyon, 2011; Park & Oliver, 2008; Van Driel & Berry, 2012).

Hutner and Markman (2016) tackle what appears to be the somewhat idiosyncratic use of beliefs by science teachers in classroom practice by proposing an operational definition of science teachers' beliefs. They define a belief as "a mental representation that influences the practice of a teacher if and only if the belief is active in cognition" (Hutner & Markman, 2016, p. 675).

An important set of beliefs that contribute to teachers' conceptions of summative assessment are beliefs about ethical practice and fairness. Because summative assessment is such an influential tool, it is important that ethical consideration is given to its use and effects (Gipps, 1994; Popham, 2000). Research has found that beliefs about ethical practice in summative assessment can vary considerably between teachers. Green, Johnson, Kim, and Pope (2007), working in the USA, for example, found that teachers generally agreed about ethical courses of action in areas of communication about grading, confidentiality and multiple assessment opportunities, but highly disagreed on standardised testing administration, bias and grading practices, so much so that the authors suggested that assessment is an area without professional consensus.

In her recent review of fairness in classroom assessment, Tierney (2013) found multiple and conflicting interpretations of fairness in the research literature. For

example, in a Canadian study, teachers' interpretations of fairness were found to vary considerably and were contradictory (Tierney et al., 2011). Teachers in this Canadian study repeatedly referred to 'being fair' in their grading practice, but the researchers found that their decisions were sometimes directed by unexamined assumptions; for example, assuming a student has not learnt because their behaviour had been poor, or that they may have copied a neighbour's work, rather than teacher decisions being directed by a comprehensive understanding of grading principles. It was concluded that given teachers could provide a clear rationale for the decisions they made it was likely that "teachers' assessment decisions may be guided more by practical knowledge and moral beliefs than theoretical knowledge" (Tierney, 2013, p. 137).

In summary, teachers' conceptions of summative assessment influence their teaching and assessment practice. Their conceptions are idiosyncratic, as they are linked to their personal beliefs and background experiences and have developed over time. Little research has been focused on important elements such as ethics and fairness in assessment, but researchers working in this field call for further professional development to be made available for preservice and inservice teachers to enable them to examine their conceptions in order to see their impact on their assessment literacy.

3.1.3.3   *The effect of emotion*

McLeod's (1992) review of research on affective constructs distinguishes between beliefs, attitudes and emotions by considering the degree to which cognition plays a role in the response, and in the time that they take to develop:

> Therefore, we can think of beliefs, attitudes, and emotions as representing increasing levels of affective involvement, decreasing levels of cognitive involvement, increasing levels of intensity of response, and decreasing levels of response stability. (McLeod, 1992, pp. 578–579)

The impact of emotion on teacher development is important to consider as emotion has been shown to affect teachers' cognition, motivation and their approaches to teaching (Sutton & Wheatley, 2003). It has also been argued that emotions are inseparable from teaching and learning science (Kayumova & Tippins, 2016; Siry & Brendel, 2016; Zembylas, 2014). The influence of emotion

in teacher development is evident through the work of Stuart and Thurlow (2000), who found that preservice teachers recalled the emotion connected with their own specific school experiences as students, and that this shaped their views of teaching practice. Furthermore, in their study of 19 preservice teachers in Finland, Anttila, Pyhältö, Soini, and Pietarinen (2016) found that both academic work and teaching practice generated a range of emotions for the preservice teachers, both positive and negative. The complex dynamic between the preservice teachers and their learning contexts appeared to regulate their academic emotions. The findings of this study drew attention to the need for preservice teachers to identify and regulate emotions to safeguard their own well-being, as well as to benefit their learning environment, accepting that "becoming a teacher is an emotionally-charged journey" (Anttila et al., 2016, p. 466). Consequently, it is important that teacher educators make preservice teachers aware of the emotional nature of their work and their own emotional responses to it (Powietrzyńska & Gangji, 2016).

Of particular note is the effect of relationships in the emotional work of teachers. Teachers' emotions are typically evoked by interactions with those in the teaching environment (Bellocchi et al., 2014; Day & Gu, 2014; Rinchen, Ritchie, & Bellocchi, 2016), and so not surprisingly student-teacher relationships have been found to be an important trigger for emotions (Hagenauer, Hascher, & Volet, 2015; Zembylas, 2002, 2004). In their study of Austrian teachers, Hagenauer, Hascher and Votet (2015) found that interpersonal relationships formed between teachers and students were the strongest predictors of joy and anxiety for teachers, and lack of discipline was a strong predictor of anger. Teacher-student relationships played a strong role in teachers' emotional experiences in class. Likewise, Hargreaves (2005) contended that teachers' emotional bonds with students could affect their orientations and actions, including their responses to change. The emotional work of caring for students has been documented for beginner teachers in a number of studies and this work has been found to be personally demanding (Bell, 2011; Hargreaves, 2000; O'Connor, 2008; Weinstein, 1998). Consequently it is worth considering what emotion could play in summative assessment literacy.

Teachers' emotions are enacted in the context of teaching and learning, and "issues about content, curriculum and pedagogy cannot be separated from

emotional issues and that all these are inseparable to a teacher's PCK" (Zembylas, 2007, p. 356). Zembylas (2007) contends that teacher knowledge includes the emotional knowledge that allows teachers to connect with the students, the content, themselves and curricular development. Studies have found that emotional knowledge and energy is used as teachers connect with subject matter, students, curriculum planning and their actions (Isenbarger & Zembylas, 2006; Zembylas, 2004; Zembylas & Barker, 2002). The frame of reference of teacher knowledge then moves from simply being within individual teachers to include wider relational and socio-political frames.

Points of transition can be influential and potentially emotionally charged, especially as beginner teachers complete their ITE programmes and move into their first jobs. At this time teachers are in a vulnerable position, as the transitions into work are often abrupt and can be traumatic (Haggarty & Postlethwaite, 2012; Huberman, 1991; Lortie, 1975; Vonk, 1993). Studies in a range of countries have shown that novice science teachers experience intense emotions in their teaching, and feel passionate about their role both with their students and amongst colleagues, and for teaching science (Nias, 1989; Ritchie, Tobin, Hudson, Roth, & Mergard, 2011; Wilson & Demetriou, 2007; Winograd, 2003).

Beginner teachers' affective experiences in their first schools are critical to their future satisfaction and success. In particular, the induction and support that beginner teachers get from mentor teachers is influential (Adoniou, 2013; Grudnoff, 2007; Haigh & Anthony, 2012). In her investigation of the transition from preservice teacher to teacher, Grudnoff (2007) found that professional and social relationships within their schools affected the teachers' abilities to become successful.

In summary, emotion is an influential element in teacher development. As discussed in Section 2.1.5, consequences of summative assessment include emotions for both students and teachers. Therefore, it is likely that as preservice teachers develop summative assessment literacy they will encounter emotional reactions, whether it be their own, or from other sources. For that reason, the effects of emotions should be considered when investigating the development of preservice teacher summative assessment literacy in this study.

3.1.4 Teacher development over time

'Becoming a teacher' infers a professional developmental journey that takes place over time and indicates some sort of trajectory. Although end points have been described in the form of graduating teacher standards and registered teacher criteria in New Zealand, these criteria also insist on ongoing growth and development throughout the career of a teacher. This section considers literature regarding teacher development over time.

Teacher professional growth has been conceptualised as stage theory by a number of writers. For example, Vonk (1993) identified two distinct stages in beginning teachers' professional development: Threshold, and Growing Into the Profession. This stage theory view matches a cognition-only view of learning. In the first year of teachers' work they are likely to be in the threshold stage, as teachers are confronted with the reality of the demands of full role of the teacher, called "reality shock" by Veenman (1984). As they get used to the role, teachers move into the second or "growing" phase.

Although their work did not focus specifically on teachers, Dreyfus and Dreyfus (1986) proposed that five developmental stages are evident as a person acquires skills through instruction and experience. These were designated novice, competence, proficiency, expertise and mastery and reflect what happens over time when students become less dependent on abstract principles and more on concrete experience. In his study investigating the development of assessment expertise in science teachers, Lyon (2013a) used the framework suggested by Dreyfus (2004) and Dreyfus and Dreyfus (1986) to conceptualise teacher assessment expertise development over time. He employed the ideas about the shift from 'knowing that', which is based on adhering to generally accepted rules and principles, to 'knowing how', which calls on the more "flexible application of principles in practice to adapt to constantly changing situations" (Lyon, 2013a, p. 1210). This led to his development of a rubric to measure the development of assessment expertise. Thus

> this relationship can vary on a spectrum from novice (limited understanding of concepts) to advanced beginner (with a global understanding, but inflexible use of concepts) to analytical decision maker (with an organized plan that draws on their understanding) to

an expertise (using concepts flexibly, depending on the curricular and student context). (Lyon, 2013c, p. 3)

So some researchers have found that the concept of stages has been found useful to describe developmental trajectories of teachers within the broader picture of their learning context (Lyon, 2013a; Schneider & Plasman, 2011). However, whether the professional learning of beginner teachers occurs in stages and is sequential, or differs for individuals and depends on context, is debated in the literature with a number of scholars rejecting the notion that development is linear and unaffected by school context (Friedrichsen & Berry, 2015; Bianchini, 2012; Renwick & Vise, 1993). Pigge and Marso (1997) conclude that "the development of the teaching concerns may not follow a lock stepped pattern, but may vary for individuals" (p. 234). It has been found, for example, in a study which investigated preservice teachers' values and practice in relation to assessment, that a simple staged model of development was inadequate to explain the differences in teachers' development (Winterbottom, Taber et al., 2008).

Rather than drawing on a stage model, Flores (2003) emphasised the situated nature of teachers' learning and professional knowledge, which is conceptualised as being socially constructed by individuals with others. She considered that beginner teachers draw on their past experiences and prior knowledge and those of others in their community as they construct their teacher knowledge in their first two years of teaching. The development of expert knowledge through conscious deliberation and reflection has been described as "theorizing practical knowledge and practicalizing theoretical knowledge" (Tsui, 2009, p. 429). This allows for the development of a more flexible and responsive application of professional knowledge.

As discussed in Section 3.1.1, teachers' learning, development and growth over time has been theorised by a number of researchers focusing on both individual and contextual influences. Perspectives on teacher learning and development over time, (see Bell & Gilbert, 1994; Clarke & Hollingsworth, 2002; Carless, 2005; Cameron, Mulholland & Branson, 2013) acknowledge the "complex web of relationships between teacher-learners and their learning" (Cameron et al., 2013, p. 12), which are not explained by linear relationships. These perspectives emphasise the complexity of teacher learning and growth by indicating

interactions between domains or levels that occur in non-linear ways, and processes of reflection and action allow change, development and growth to occur over time in work contexts. Non-linear perspectives allow for idiosyncratic pathways in growth to be plotted or described, and "provide recognition of the situated and personal nature of teacher growth" (Clarke & Hollingsworth, 2002, p. 965). When discussing the personal domain, local school context domain, and the domain of wider external forces, Carless (2005) asserts that "the impact of implementation [of an assessment change process] involves a complex interplay of these, and undoubtedly other factors" (p. 51).

In case studies presented by Bullough (1992), the role of mentor teachers was highlighted as playing an important part in beginner teacher development. School contexts were also found to be powerful in limiting or enabling teachers to develop their role over time. Given the strong influences of past experience, personality and context, the conclusion was drawn that

> becoming a teacher is inevitably an idiosyncratic process … but the personal must be considered in relationship to the contextual, one without the other represents a lopsided education and an extremely naïve view of human development and the power of institutions to influence development. (Bullough, 1992, p. 251)

Bullough (1992) highlighted four areas of knowledge as affecting beginning teachers' development over time: PCK, knowledge of the relationship between curriculum decision making and teacher roles, knowledge of self, and knowledge about beginning teacher development. It is argued that in ITE some teachers believe that they have sufficient knowledge by the end of their teacher education programme, leading them to believe they are fully equipped to teach (Wilson & Demetriou, 2007). However, teachers need to continue to focus on developing knowledge and understanding so as to further their professional growth, as there exists an inherent difficulty in transferring knowledge from one context to another or bridging from one role to another (Bianchini, 2012; Kang et al., 2012). Assisting beginner teachers with these challenges is the role of those working in ITE and those working as mentor teachers in schools.

Previous studies do show that many teachers (which would include those working in the role of mentor teachers) are inadequately prepared and are challenged by

aspects of assessment (Bol & Strage, 1996; Hill, 2000; Izci, 2013; Koh, 2011, 2014; Suurtamm & Koch, 2014). Gaps in assessment literacy in theory and practice can exist for even experienced teachers (James & Pedder, 2006; Korthagen, 2010; Ogan-Bekiroglu & Suzuk, 2014), and it has been suggested that teachers need opportunities to reflect, practise and revise their assessment methods (Simon & Campbell, 2012). Professional development conducted over a longer time frame has been shown to help teachers develop assessment literacy (Koh, 2011; Poskitt & Taylor, 2008).

Few studies have tracked teacher development through ITE and into early career. In one such study, science teachers were shown to find difficulty as they changed from student to teacher, particularly in bridging the gap between science knowledge and pedagogical knowledge (Kang, Bianchini, & Kelly, 2012). One four-year longitudinal study (Grossman et al., 2000) suggests that teachers do draw on the pedagogical tools introduced during their ITE. However, the study found that these tools were not always evident in the first year of practice but that they became more visible in the second year once teachers were more able to focus on good teaching of a subject area. Earlier, similar findings were reported by Trumbell (1999), who found that by their third year teachers showed that they were more reflective and critical of their practice. In another four-year study, which focused on teachers' use of assessment tools, it was found that assessment tools acted as boundary objects as teachers moved from university courses to practicum, and then to teaching positions (Nolen, Horn, Ward, & Childers, 2011). A feature of these studies is that the contexts in which the teachers were operating were shown to be influential in the teachers' development of understanding and practices.

In summary, research has shown that a number of influences are present as teachers develop their knowledge and associated practice in a number of contexts over time. These influences come from a number of sources, both external and internal to the teacher. Non-linear perspectives allow for idiosyncratic trajectories to be considered, with combinations of influences found to affect teachers in a various ways.

In the next section I present two models that will help frame this study: a model for science teacher assessment literacy developed by Abell and Siegel (2011), and a model of teacher professional knowledge and skills (Gess-Newsome, 2015). These will provide a useful structure for reflecting on science teachers' knowledge and action, and their use enables an informed consideration of the development of summative assessment literacy over time.

3.2 Models which include teacher knowledge with a focus on knowledge for science assessment

The strong link that exists between teaching, assessment and curriculum means that science teachers' summative assessment literacy cannot be wholly separated from their understanding of the concepts and practices of science nor from the science teaching context. Science requires students to learn about complex natural phenomena for which they may already hold alternative conceptions, as well as learning about the nature of science and scientific inquiry (Cowie, 2013; Duschl, 2008). Science teachers need to help students to understand and be able to apply the scientific method within theoretical and practical investigations, which can be complex and problematic as it calls on particular knowledge and skill sets (Loughran et al., 2006). Not surprisingly then, experienced science teachers' actions, including those relating to assessment, have been shown to be directed and justified by the unique set of knowledge and skills that are part of science education (Rozenszajn & Yarden, 2014). Assessment in science education has already been the subject of considerable research and review, which illustrates the challenges teachers face in assessing students' science learning (Abell & Siegel, 2011; Bell, 2007; Britton & Schneider, 2007; Corrigan et al., 2013; Cowie, 2013; Edwards, 2013a, 2013b). The conception of PCK (Shulman, 1986) has been of particular influence on research in science education and science teacher practice.

Shulman (1986) introduced the concept of a specialised form of knowledge developed by teachers and used for teaching: PCK. He described PCK as the knowledge that is developed by teachers to help others learn. It is usually understood as an amalgam or blending of content and pedagogy, where the teacher combines their understanding of the content they want to teach with their knowledge of instructional strategies and the specific student group in order to help students to learn that content (Kind, 2009). In Shulman's terms it is

that special amalgam of content and pedagogy that is uniquely the province of teachers…PCK…represents the blending of content and pedagogy into an understanding of how particular topics, problems or issues are organized, represented, and adapted to diverse interests and abilities of learners, and presented for instruction" (Shulman, 1987, p. 8).

PCK is influenced by and influences the other knowledge bases teachers use for teaching (Magnusson et al., 1999). In science teaching, for example, PCK is influenced by science subject knowledge, pedagogical knowledge and knowledge of context (Grossman, 1990; Shulman, 1986). Magnusson et al. (1999) describe PCK as the result of the transformation of knowledge from other domains, namely subject matter knowledge and beliefs, pedagogical knowledge and beliefs, and knowledge and beliefs about context. They also note that PCK in turn can affect development of these knowledge bases.

There have been a number of models proposed to conceptualise the construct of PCK, and the relationships between its components and contributing elements. The most cited in science education is a model developed by Magnusson, Krajick, and Borko (1999). Within this model assessment knowledge is seen as one of the components of PCK for science teaching (Grossman, 1990; Magnusson et al., 1999; Tamir, 1988). A range of other models have since been published, some based on the work of Magnusson et al. (1999) (e.g., those developed by Abell & Siegel, 2011; Ball, Thames, & Phelps, 2008; Borowski, Carlson, Fischer, Henze, & Gess-Newsome, 2012; Gess-Newsome, 2015; Gess-Newsome & Lederman, 1999; Meis Friedrichsen & Dana, 2005; Park & Oliver, 2008). Loughran et al. (2006) remarked that PCK is not a single entity identical for all teachers of a given subject, but that it is an expertise that is idiosyncratic and influenced by the teacher's experience, the teaching context, and the content; "it is, nevertheless, a cornerstone of teachers' professional knowledge and expertise" (p. 9). Many of the models published include assessment knowledge within PCK either explicitly or implicitly, but few focus on assessment literacy. The remainder of this section focuses on two models that have been developed recently to explain aspects of teacher knowledge and development.

The first model is a model for science teacher assessment literacy (Figure 2), developed by Abell and Siegel (2011). The model is based on the authors' own

empirical work, and built on the theoretical foundations provided by Magnusson et al. (1999), and Pellegrino et al. (2001), who defined assessment as a process of reasoning from evidence and provided a useful framework. According to Abell and Siegel (2011), their model "attempts to capture various types of assessment knowledge and skills that we believe teachers need to create an assessment centred learning environment" (p. 211).

Knowledge of Assessment Purposes	Knowledge of what to assess
Assessment Values Views of Learning **And Principles**	
Knowledge of assessment strategies	Knowledge of assessment interpretation and action-taking
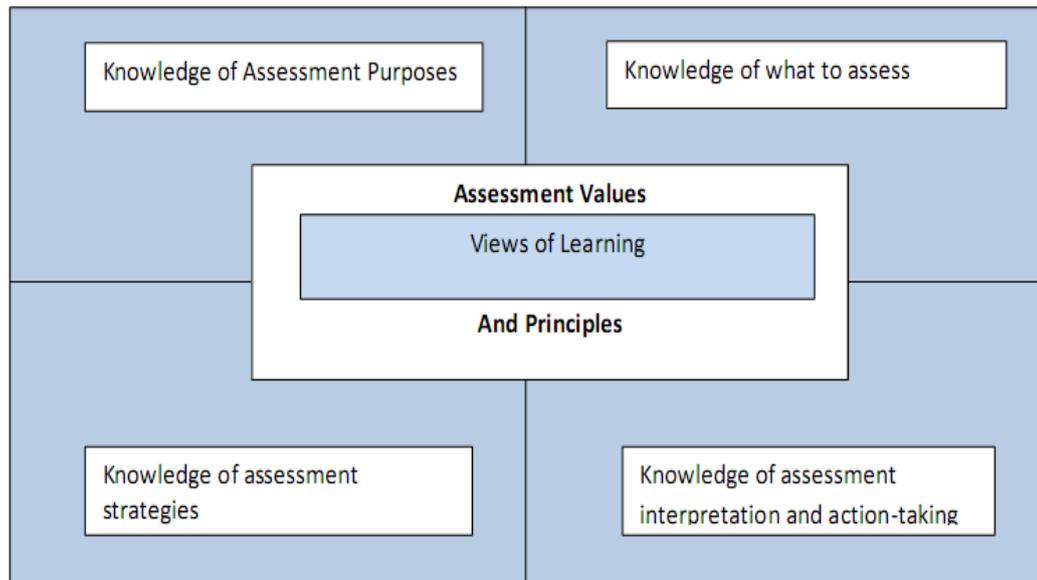
Figure 2.   A model for science teacher assessment literacy (Abell & Siegel, 2011, p. 212)

At the core of this model are teachers' views of learning and the assessment values and principles which guide their assessment decisions. Abell and Siegel (2011) comment on the central and pivotal role of views of learning, values and principles in guiding the practice of assessment. Four principles of effective assessment are listed as particularly important: assessment is a process through which teachers can learn; assessment is a process through which students should learn; assessment should help students be metacognitively aware of their developing knowledge and skills so they can self-regulate their learning; and assessment needs to be equitable for all learners. The model proposes that the values and principles a teacher holds form the core of their assessment literacy. The values and principles then interact with the four categories of science teacher knowledge of assessment (which also interact with each other): assessment purposes, what to assess, assessment strategies, and assessment interpretation and action taking. Although the four kinds of knowledge seem to be

compartmentalised in the model, Abell and Siegel demonstrate that teachers connect these elements of knowledge in their assessment practices.
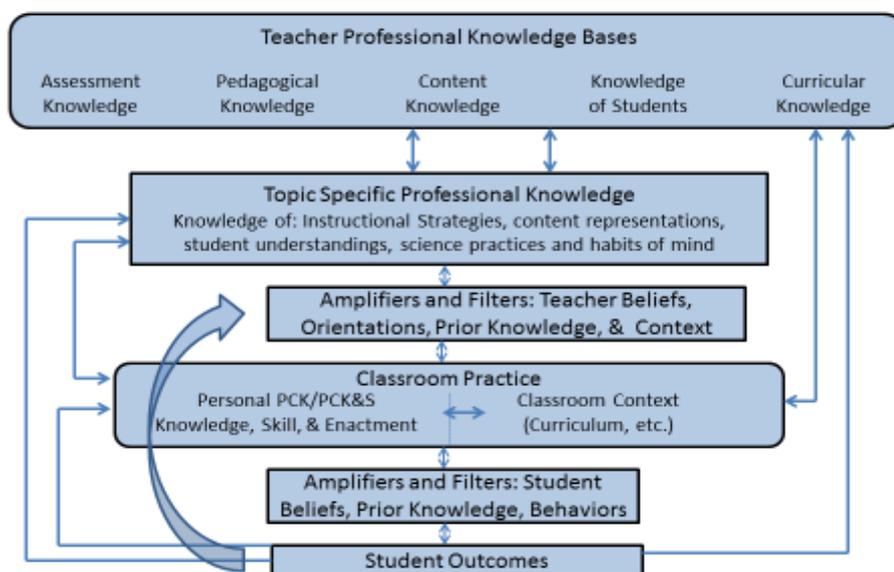
Figure 3. Model of teacher professional knowledge and skill (Gess-Newsome, 2015, p. 31)

The second model is the Model of Teacher Professional Knowledge and Skills (TPK&S) (Gess-Newsome, 2015), Figure 3. This model was the outcome of a summit held by key researchers in the area of PCK in 2012 to further theorise PCK. Shulman identified four limitations in earlier formulations of PCK: the lack of non-cognitive attributes such as emotion, affect, moral judgment and reasoning in teaching; the focus on the intellectual and insufficient attention given to pedagogical action; the lack of attention to the cultural and contextual elements and the situated nature of teaching; and the lack of focus on outcomes (Shulman, 2015). This model addresses Shulman's concerns by unpacking many of the ideas previously packed into PCK. The relationships between teacher knowledges and action are made clear, with both teacher professional knowledge bases and topic specific knowledge conceptualised as generic or public knowledge, therefore context free. Non-cognitive attributes such as affect and emotion, as well as teachers' beliefs, orientations, and cultural and contextual elements are acknowledged as acting as amplifiers or filters to teacher learning and practice as they personalise knowledge. This model acknowledges the role of classroom context and its interactions with teachers' personal PCK. Finally the model

includes a focus on student outcomes, and acknowledges the feedback loops that can result from these outcomes.

The Model of Teacher Professional Knowledge and Skills (TPK&S) identifies the overarching role of teacher professional knowledge bases and suggests that when teaching, teachers draw from a number of knowledge bases: assessment knowledge, pedagogical knowledge, content knowledge, knowledge of students, curricular knowledge. The teacher professional knowledge bases inform and are informed by topic specific professional knowledge, that is, knowledge which encompasses what is required to teach specific science topics. For example, if teaching students about photosynthesis, teachers will call on knowledge of proven strategies for instruction of this topic, content representations, common alternative conceptions held by students about photosynthesis, and so on. In this model both professional knowledge bases and topic specific knowledge are conceptualised as context free, normative, expert knowledge as understood by the academic community.

However, the TPK&S model recognises the personalisation of these knowledges by teachers as it passes "through the lens of the teacher" (Gess-Newsome, 2015, p. 34). Teachers' views, beliefs, experiences, affective responses and the context are considered to act as amplifiers and filters, influencing individual teachers' knowledge and development as well as their classroom decision making and resulting practice. These amplifiers and filters mediate the translation of teacher professional knowledge bases to classroom practice "as interactions between knowledge, affect, and practice are dynamic, the combination may create unpredictable and idiosyncratic implications for different teachers" (Gess-Newsome, 2015, p. 35). Consequently what teachers prioritise and pay attention to when making classroom and assessment decisions is understood as idiosyncratic because it is influenced by their individual set of amplifiers and filters.

As an example of amplifiers and filters in action, Gess-Newsome (2015) proposes that a teacher who believes that teaching is telling might reject constructivist inspired pedagogy. In this case the teacher may have been taught about constructivist views of learning and associated instructional strategies, but his

own belief will override suggestions that the strategies will be useful when teaching (Gess-Newsome, 2015). Because of his beliefs, such a teacher may be more motivated to learn about how to set up multiple choice quizzes to assess factual knowledge that he has communicated/ transmitted than to learn how to assess how students develop ideas about a topic through a portfolio of work. If however his school implements policy to broaden the assessment tools used, he may be required to assess in ways in which he does not wholeheartedly agree. Corrigan et al. (2013) similarly identify filters in action for teachers as they interact with assessment and curriculum frameworks within a policy setting and develop their pedagogy and PCK. Their "knowledge of worth filter" (Corrigan et al., 2013, p. 3) is proposed to be composed of school and teacher values, and beliefs about science, about science education in a particular cultural setting, and about assessment. A number of studies have shown the 'filtering' effects of beliefs on teachers' practice, but also the impacts of school culture which may challenge their beliefs (Askew, 2013; Corrigan & Cooper, 2013; Ma, 2013).

The models presented by Abell and Siegel (2011) and Gess-Newsome (2015) have a number of elements in common. They both acknowledge the importance of a knowledge base for assessment and teacher action, and the variation in how this knowledge base is developed and utilised depending on the lens of the teacher and the nature of the context in which the teacher is working. They each put forward an "amalgam" of different knowledges as foundational to summative assessment literacy. They both have been derived from research that focuses on science teaching, which means they reflect more readily on what and how science teachers prioritise and why they might decide to act, or not. Chapter 2 has illustrated and argued that summative assessment literacy for science teachers includes their understanding and application of generic assessment concepts, and also topic-specific assessment knowledge and skills.

A number of science education researchers argue that consideration of PCK is a useful starting point when conceptualising how teachers might develop assessment expertise (Lyon, 2013a; Wang, Wang, & Huang, 2008). In his study, which focuses on the development of assessment expertise in science teachers, Lyon (2013a) argues that the construct of PCK helps researchers understand that "science teachers' knowledge of assessment is not relegated to domain-

independent strategies, but rather entails considering what and how to assess in *science* classrooms" (Lyon, 2013a, p. 1209, emphasis in original). This idea is similarly emphasised by Steiner Engelsen and Smith (2014) who argue that the practice of Assessment for Learning (AfL) is embodied in teachers' PCK, and by Jones and Moreland (2005) who found that teachers with weak PCK did not engage readily with students in assessment conversations and did not provide enough relevant feedback to students about their learning. For the purposes of this thesis both models are utilised.

The model of science teacher assessment literacy developed by Abell and Siegel (2011), used alongside the model of TPK&S proposed by Gess-Newsome (2015), has been chosen for use in this study. Together these provide a useful structure for reflecting on science teachers' knowledge bases for action and the wider influences on their decision making as this relates to summative assessment. In the study they are used to enable an informed consideration of the development of summative assessment literacy for beginner science teachers.

3.3 The development of teachers' assessment literacy

Assessment literacy research is a relatively new field, and summative assessment literacy has not often been written about separately but is instead usually included in the literature regarding assessment literacy. Most of the literature is focused on the description of the features of assessment literacy or how it can be measured (Bakx, Baartman, & van Schilt-Mol, 2014; Chen, 2005; Christoforidou, Kyriakides, Antoniou, & Creemers, 2014; Mertler & Campbell, 2005; Schneider & Plasman, 2011). Far less literature focuses on how assessment literacy develops over time for teachers and what facilitates the development of assessment literacy. Previous studies have shown that preservice teachers graduate with a limited understanding and appreciation of classroom assessment (DeLuca & Klinger, 2010; Hill et al., 2013; Maclellan, 2004; Popham, 2009) but research into early stages of assessment literacy development is scarce (Abell, 2007). The following is a review of literature that focuses on the development of assessment literacy for teachers. It highlights the influences on assessment literacy development of ITE and practicum experiences as well as mentors and describes relevant studies

within the field of science education. Research focusing on the development of assessment literacy in the New Zealand context follows in Section 3.4.

ITE programmes are identified as a key element in the preparation of beginner teachers but research findings on the role of ITE in developing assessment literacy are inconsistent, with studies showing that some ITE programmes make a more significant difference for teachers' assessment literacy than others (Chen, 2005; Cooper & Edwards, 2013; DeLuca & Klinger, 2010; DeLuca, Chavez, & Cao, 2013; Eyers, 2014; Maclellan, 2004; Ogan-Bekiroglu & Suzuk, 2014; Volante & Fazio, 2007). Some ITE programmes prescribe specific assessment courses whereas others embed assessment education in a number of their courses. Although some research has shown the benefit of integrated ITE courses which embed assessment (e.g., DeLuca, Ogden, & Pero, 2015), many researchers call for the inclusion of preservice courses which teach assessment explicitly, as these have been shown to make measureable change in teachers' understandings and practice, albeit less consistently for teachers than would be hoped (Alkharusi, Kazem, & Al-Musawai, 2011; Chen, 2005; DeLuca & Klinger, 2010; Smith et al., 2014).

There is evidence that ITE programmes that include explicit assessment education courses enable preservice teachers to develop a foundation of understanding from which to continue learning about assessment (DeLuca, Klinger et al., 2010; DeLuca, Chavez, Bellara et al., 2013; DeLuca, Chavez, & Cao, 2013; Eyers, 2014). For example, a study based in the USA, exploring preservice teachers' confidence levels, investigated the development of confidence and conceptions in preservice teachers (DeLuca, Chavez, & Cao, 2013). By comparing data before and after an assessment course it was found that preservice teachers showed evidence of developed confidence and deepened conceptions of assessment, alongside a move from simplistic to complex understandings of assessment. This study supported the use of courses that explicitly teach assessment within an ITE programme. Similar conclusions were drawn by McGee and Colby (2014), who also found that a compulsory assessment course did serve as a foundation for future learning about assessment.

While McGee and Colby (2014) did not find any differences in outcomes for early years, primary and secondary student teachers, DeLuca and Klinger (2010), working in Canada, found that the involvement in an additional elective assessment course significantly developed primary/junior preservice teachers' confidence levels and perceived readiness, in contrast with intermediate/secondary preservice teachers. This study found that preservice teachers held relatively high levels of confidence after an elective course, but on closer inspection it was found that the teachers were more confident in tasks related to summative assessment than formative assessment. DeLuca and Klinger (2010) suggest this could be because the preservice teachers had more exposure to summative assessment during their own education. However, the authors caution that teacher candidates' high confidence ratings may represent a level of over-confidence, given the complexity of technical issues associated with summative assessment. This study highlighted the benefits for teacher candidates of direct instruction on assessment to complement the learning about assessment that occurs in other aspects of ITE programmes.

There are very few research studies which examine assessment literacy development in preservice teachers over a time frame longer than a year. Research by Volante and Fazio (2007) in Canada is one such study. These researchers investigated the development of assessment literacy during a four year primary ITE programme. Their study found that the preservice teachers retained low levels of self-efficacy with respect to assessment over the four year period and that they mostly referred to summative assessment. Preservice teachers in the study overwhelmingly endorsed the use of specific courses within their ITE to help develop their assessment literacy. A key finding from the study was the impact of mentor teachers during preservice teachers' practicum experiences. Mentorship during practicum through appropriate modelling and the provision of feedback was identified by participants as important for their development, but was not always present. These findings emphasising the importance of the roles of mentor teachers echo the findings in other studies on preservice teacher understanding of assessment (DeLuca & Klinger, 2010; Graham, 2005; Gustafson, Guilbert, & MacDonald, 2002; Luft, 2009; Mueller & Skamp, 2003).

On narrowing the focus to research studies investigating the nature of science teachers' assessment literacy, very few studies exist on the development of assessment literacy in science teachers. Given the dominance of constructivist views of learning, and the variety of knowledge and skills that are assessed in science, the assessment literacy required to assess science is substantial. In one study of the development of PCK in science, of which knowledge of the assessment was identified as a part, Park and Oliver (2008) provided case studies of three teachers. Their study found that PCK development occurred as a result of teacher reflection, and its development was linked to the efficacy of the teachers involved. They also found that students played a role in the way PCK was organised and developed and the teachers' understandings of students' misconceptions had a major impact on assessment PCK. PCK development was strongly influenced by teachers' understandings of students' misconceptions. The study also drew attention to the idiosyncratic nature of PCK enactments. Importantly, this study drew attention to the interrelationship of knowledge acquisition and knowledge use for PCK development, as teachers acquired knowledge through their practice. It was found that the most powerful changes to teachers PCK occurred through teachers' practical experiences. This study also raised the issue of an affective component of PCK: teacher efficacy. Park and Oliver (2008) argue that teacher efficacy is linked with PCK and that teacher efficacy serves as a conduit connecting the two dimensions of PCK: understanding and enactment.

In a study involving 215 Taiwanese preservice elementary teachers, focusing on science assessment, Wang, Kao, and Lin (2010) found that teachers' mode of assessment was aligned with a traditional view of learning (empiricist tendency), not with the constructivist approach they were learning as part of their ITE. Their understanding of assessment probably reflected the traditional school culture in Taiwan. Beginner teachers were found to focus on assessments as a means to assign grades to students, to motivate students to work and see whether they needed to repeat the teaching of material (Wang et al., 2010). As teachers developed they were found to focus more on the formative uses of assessment in order to provide feedback to students, and then to find out about students' thinking about specific science ideas. This study suggested the need for ongoing

96

attention to and reflection on the coherence between learning, curriculum and assessment in ITE programmes in order to help preservice teachers reconstruct their assessment practices. Similarly, Siegel and Wissehr (2011) found that although preservice secondary science teachers recognised assessment tools and the need to align the assessment tools they used with their learning goals, they tended to revert to traditional forms of assessment in their science units. These authors suggest that more emphasis on the development of assessment literacy during ITE is needed.

In their review of science teachers' PCK development, Schneider and Plasman (2011) developed progressions that reflected increasingly sophisticated knowledge, including PCK related to the assessment of students' science learning. Two categories were identified as relevant to this: "teachers' ideas about *strategies* for assessing student thinking in science" and "how or when to *use* science assessments" (Schneider & Plasman, 2011, p. 538, emphasis in original). Firstly, teachers' ideas about strategies for assessing were found typically to be absent from the planning of beginner teachers and assessment tended to be in traditional formats (e.g., tests). Teachers' development of assessment PCK was seen to begin when they included informal questioning and a wider variety of strategies such as journal entries, portfolios and so on in their day-to-day practice (Goodnough & Hung, 2009). Experienced teachers were found to have a better understanding of the links between science learning and assessment, with evidence in the form of assessment task design knowledge (Gearhart & Osmundson, 2009). This meant that they understood the relationship between the science and the assessment tasks. Gearhart et al. (2006) found that the interpretation of students' work enabled teachers to develop a more sophisticated understanding of assessment practice over time. This suggests that experience in working with students is necessary to assist beginner teachers to think more carefully about strategies for assessment in science.

The second of the two aspects of development described by Schneider and Plasman (2011) was the knowledge of how or when to use science assessments. As expertise and knowledge developed further, Schneider and Plasman (2011) found that the research showed that teachers started to link students' responses in assessments to their own practice in order to improve or adjust instruction. Again,

for the development of expertise with respect to how and when to use science assessments, experiences with students seemed to be a very important factor. Alongside the experience with students, guidance from formal courses "pushed teachers to think about how to assess at higher levels of thinking" (Schneider & Plasman, 2011, p. 555).

Lyon (2012, 2013a, 2013b, 2013c) investigated the development of USA preservice science teachers' assessment expertise with respect to construction of assessment tasks, use of assessment tasks and understanding about equity in assessment. He developed a tool to measure or track the development or shifts in assessment expertise. Over a one-year programme, development in all three aspects of assessment expertise were measured and preservice teachers were found to gain expertise through their ITE programme. They developed an increased understanding of the need for coherence between learning objectives and assessment tasks, but were still developing their understanding of links between tasks, learning objectives and evaluative criteria. Over time, the teachers in Lyon's study expressed higher capacity to modify assessment materials but not to develop their own materials from scratch. Teachers were found to move towards a more formative view of assessment. The preservice teachers in Lyon's study also showed movement towards a more situated view of assessment, increasingly demonstrating their ability to not only know about assessment but also to know how to assess within the context in which they were working. With respect to equity in assessment, the preservice teachers in Lyon's study demonstrated more awareness of pertinent issues, but this was rarely evident in their teaching plans.

Individual case studies of preservice teachers involved in Lyon's study (Lyon, 2013c) show that their developmental trajectories were idiosyncratic. Factors affecting their development included their interactions with a diversity of students in practicum schools that allowed them more or less experience with English language learners, as this fed into their learning about equity in assessment. The actual topics they were able to teach on practicum schools also provided them with varying experiences in assessment. Their university teacher had an effect on their development by way of the instruction he offered, especially in relation to equity issues for English language learners. Lyon's study indicated that science

classroom assessment is complex, and the development of expertise requires considerable effort. His findings emphasised the need to support inexperienced teachers to assess science in ways that are equitable and that promote academic excellence (Lyon, 2012).

In summary, the development of teachers' assessment literacy is influenced by ITE experiences and practical experiences, as well as their own conceptions (Graham, 2005; Levin & He, 2008; Nolen, Horn, Ward, & Childers, 2011; Struyven, Dochy, & Janssens, 2008; Volante & Fazio, 2007). The use of compulsory or elective courses has shown to promote the development of assessment literacy for preservice teachers and provide a framework which teachers can use in their development. However, preservice teachers do not necessarily respond to these courses in predictable ways. Some literature argues for a more integrated approach to assessment education. Overall, the research into the ways ITE can best help preservice teachers develop assessment literacy points to the difficulty and complexity of the process. Practical experience in working with students is an important contributor to beginner teachers' learning, and the assessment literacy of ATs and mentors may help or hinder the process of assessment literacy development for beginner teachers.

The next section reviews research specifically related to the development of assessment literacy in the New Zealand context.

3.4 The development of teacher assessment literacy in the New Zealand context

As described in Section 2.2.1, New Zealand secondary school teachers work in an environment in which they have considerable responsibility as assessors, including assessing students against standards for school-leaver qualifications. The context is one of high trust, where the professionalism of teachers is expected and required. A good understanding and practice of summative assessment is therefore essential to enable practice when newly graduated teachers start teaching.

Little research on the development of assessment literacy in beginner teachers has been conducted in New Zealand and there have been calls for further work in this area (Hill et al., 2010). A recent study across four universities in New Zealand

investigated the development of assessment capability in preservice primary and early childhood teachers during their three year undergraduate ITE programmes (Smith et al., 2014). This study found that teachers graduating from the programmes had shifted in their understanding of assessment and considered themselves ready to begin using formative assessment. However, these teachers were deemed as still emergent when their understandings were compared to the vision for assessment held for the schooling sector (Ministry of Education, 2010). The study found that high profile national policy requirements (e.g., the introduction of national standards) could tend to dominate learning about assessment, meaning lesser exposure to other equally important but perhaps lower profile policy initiatives (e.g., curriculum based priorities). The importance of relevant practical experiences during ITE was confirmed. The study found that it was important for lecturers to take account of preservice teachers' beliefs and knowledge. Teachers graduating from the programmes were deemed to still require considerable support in their first job, in order to continue their assessment development.

A smaller qualitative doctoral study examined primary preservice teachers learning about assessment in their final year of a three year ITE programme at one New Zealand university (Eyers, 2014). This study found that teachers' beliefs, understandings and practices changed, reflecting the influences and interactions at play during their ITE. The teachers attributed much of their development to an assessment course they completed in the second year of their three year programme. They found that the alignment of educational assessment policy and classroom practice enabled them to make connections between theory and practice. Encountering dissonance also enabled learning. The teachers in Eyers' (2014) study graduated with a reasonable level of confidence except when summatively assessing students, and demonstrated emerging adaptive expertise.

I found very few studies investigating beginner secondary teachers' concerns about assessment in New Zealand. One study found that secondary teachers' sense of efficacy increased in general over the period of their ITE programme, but with respect to assessment they voiced concerns about not having enough training in creating reliable assessment tasks, in communicating assessment results and in marking assessments fairly (Smith et al., 2013). In an earlier study in which

secondary preservice teachers' attitudes and understandings of assessment were investigated it was found that over time teachers changed their view from viewing assessment from the student's perspective to viewing assessment from the teacher's perspective (Cooper & Edwards, 2013; Edwards & Cooper, 2012). Preservice teachers in this study progressed from a simplistic view of assessment to one showing more understanding of its complexity, and from a summative-only view to a view encompassing formative uses of assessment, and from individual to group assessment. The preservice teachers in the study credited their increased understanding of assessment to their practicum experiences, and to a specific course from university (an assessment course taught in the final year of their programme) as well as their own experiences of being assessed throughout the programme. By the end of their four year programme the preservice teachers provided evidence of attempts to theorise assessment practices. Although they were able to link planning, teaching and assessment, and had developed a range of assessment strategies, they voiced concerns over NCEA assessment and marking and felt they still had a lot to learn.

3.5   Summary of teacher professional knowledge and assessment literacy development

The development of summative assessment literacy is crucial for beginner teachers as they advance through their ITE programme and into their classrooms. As explained in Chapter 2, summative assessment has a number of purposes and elements that need to be considered but only very limited attention has been paid to how beginner teachers develop summative assessment literacy. This chapter has reviewed literature concerned with teacher development as viewed from both a cognitive and sociocultural perspective, with particular attention to the development of assessment literacy. The literature provides evidence of the important contributions ITE can make to beginner teacher development when the university and school focus is coherent. The chapter also highlights the important contributions of teacher affect, conceptions and background experiences as well as the influence of context on teacher learning and decision making.

Models have been presented which summarise the key influences on teachers' development and action particularly related to knowledge for teaching and assessment decision making. The models by Abell and Siegel (2011) and Gess-

Newsome (2015) have been chosen for deeper exploration and use in this study because of their links to science education. Abell and Siegel also focus on assessment literacy. Together the models highlight the need to attend to teacher knowledge, beliefs, emotion and so on as well as to the context of leaning and practice. An amalgam of teacher knowledges and its use in classroom decision making but both teacher internal and contextual factors act as amplifiers and filters in their decision making processes. In this study I use these ideas to explore the development of summative assessment literacy of a small number of beginner science teachers.

3.6 Looking across the two chapters: The research questions

There are a number of purposes for summative assessment, but the over-riding purpose that is espoused in New Zealand educational policy is that all assessment is used for the improvement of student learning and teacher teaching. Bearing this in mind, it is very important for beginner teachers in New Zealand to develop summative assessment literacy so that they are well equipped to gather assessment data, consider, analyse and use it in ways that benefit student learning and represent that learning reliably.

Summative assessment literacy for science teachers includes their understanding and application of generic assessment concepts, and also topic-specific assessment knowledge and skills. While professional knowledge relating to assessment is an important characteristic of assessment literacy (Stiggins, 1995), knowledge itself is not the sole characteristic of summative assessment literacy. Summative assessment decision making is influenced by a number of factors, both internal and external to the teacher.

Review of the literature shows that little research has been done internationally that focuses on how beginning secondary science teachers develop knowledge and capabilities in the area of summative assessment, and certainly no research in this area has been found in a New Zealand context. The kinds of learning that science teachers need to assess science are in part specific to science, and the ways in which this learning can be assessed can be considered as part of the science teachers' PCK. In the very particular context of secondary schools in New Zealand, where teachers have a direct part to play in summative assessment for

qualifications, the need for them to develop summative assessment literacy is especially pertinent. In particular, it would be useful to understand what characterises assessment literacy for this sub-group of beginner teachers, how their thinking and practices change through their ITE experiences and as they start teaching, and what contributes to the development of their summative assessment literacy. To date, most research in the field of assessment literacy focuses on the description of the teachers' conceptions, beliefs, and practices at specific points in time, through the use of survey instruments. Less research has focused on teachers' development of assessment literacy over extended periods of time.

This study aims to further our knowledge in the area of beginner teacher development of summative assessment literacy by focusing on the following three research questions:

a) What are the characteristics of the summative assessment literacy as perceived by beginner secondary science teachers?

b) In what ways do beginner secondary science teachers' summative assessment literacy develop during their formal teacher education and beginner teaching experience?

c) What do beginner secondary science teachers identify as contributors to the development of their summative assessment literacy?

The next chapter outlines methodology and methods used in the research that sets out to answer these questions.

Chapter 4: Methodology and Methods

This chapter describes the research methodology and methods used in this study. It consists of eight sections and begins by grounding the research within an interpretative qualitative methodology, based on my positioning as a researcher with respect to ontology and epistemology (Section 4.1). This is followed by details of the methods used for selecting participants (Section 4.2), collecting data (Section 4.3), analysing data (Section 4.4), ensuring the quality of the research (Section 4.5), and ethical considerations (Section 4.6). Section 4.7 offers concluding comments.

4.1   Research paradigm

In any research inquiry, it is important to locate the particular theoretical and methodological underpinning adopted for the research within the researcher's assumptions and beliefs. Burrell and Morgan (1979) discuss four sets of assumptions that underpin a researcher's conceptions of the social world within which their research takes place: assumptions of an ontological kind, which concern the reality or truth of the phenomenon under investigation; assumptions of an epistemological nature, concerning knowledge and the relationship between the knower and what can be known; assumptions concerning human nature, in particular the ways humans and their environments interact; and assumptions regarding methodology, defined as a "research strategy that translates ontological and epistemological principles into guidelines that show how research is to be conducted" (Sarantakos, 2005, p. 30). That is, the choice of a research paradigm and the methods used for any research project must be fit for purpose and reflect the ontological and epistemological views of the researcher. Burrell and Morgan's four assumptions are unpacked next to explain my beliefs and show their consistency with the methodology chosen.

4.1.1   Ontology

Ontology is concerned with reality, its structure and nature. There are a range of views of the nature of social reality. Researchers holding a realist position view reality as external and discoverable, independent of human cognisance. The existence of reality is not dependent on the knower (Cohen et al., 2011). Realists

value the scientific method and empirical evidence in research and view objectivity as an important feature in any research. An alternate view is the ontology of relativism which posits that reality is not separate from us but is constructed intersubjectively by people as they develop meanings socially and experientially. Relativists reject the notion of a detached, objective observer and instead argue that knowledge is known only through the subjective experiences of people (Creswell, 2012). Relativists view knowledge as inseparable from the knower. For this study I adopt a critical realist position (Denzin & Lincoln, 2013). This position maintains that a real world exists independent of individuals and their constructions, but that to understand this world individuals must construct meaning from their particular theoretical perspective or worldview. Therefore human beings may engage in the same reality but make different meaning of it (Beattie, 1995).

4.1.2 Epistemology

Epistemology is concerned with knowledge, its nature and forms, and how it is gained and shared (Denzin & Lincoln, 2013). Knowledge and the nature of knowledge can be viewed in a number of ways; for example, as objective and able to be transmitted in tangible form, or as subjective, essentially personal and based on the interpretations of experience (Burrell & Morgan, 1979). A critical realist ontology suggests a view of knowledge that is subjective, socially constructed and personal with a fluid and dynamic view of meaning; this view allows for multiple interpretations of reality (Crotty, 1998). The 'social constructivism', which draws on cognitive, constructivist and sociocultural theories, describes this view (Shepard, 2000). Social constructivism argues that individuals construct varied and multiple meanings of their experiences and are not able to be objective or separate themselves entirely from what they are experiencing. This results in the researcher aiming to understand the complexity of views within a situation rather than distilling 'the truth' (Creswell, 2013).

Researchers working within a social constructivist epistemology recognise that they are part of their research context and that their own background (personal, cultural, historical) shapes their interpretations of what they see and hear (Creswell, 2013). In Chapter 1 I provided relevant details of my background and research interests, as an acknowledgment of what I bring to this study as a

researcher. I regard knowledge as the result of meaning making which is constructed by the interaction between people and the world around them, allowing for multiple interpretations. This perspective focuses on the "meaning making activity of the individual mind" (Schwandt, 1994, p. 127) where meaning making is seen as being relative to the contexts where it is constructed, and human agency is integral to this process.

4.1.3 Human nature

The third of Burrell and Morgan's (1979) assumptions that underpin a researcher's conceptions of the social world concern human nature and the ways humans and their environments interact. At the two extremes human beings can be viewed as reacting and responding to their environment, or as acting with free will, intentionality and outside of any structures. In the first case, the individual could be seen to be constrained by their cultural history and social structures in ways that mean their path is predetermined to a large extent, whereas in the second case they appear to be disengaged with any other person or group, hence being able to act completely on their own. As a researcher I take a more moderate approach, similar to that of Carr (1986), believing that although humans are affected by their social history and culture, they have free will and can disengage themselves from their surroundings to an extent to exercise a degree of self-determination. That is, individuals have agency in a range of domains but are also social beings who engage with the society in which they live.

4.1.4 Research methodology

The final of Burrell and Morgan's (1979) assumptions is about the research methodology. For the purposes of this study, and consistent with my underpinning beliefs as a researcher, a qualitative interpretivist methodology was selected for the research. The interpretivist paradigm is concerned with describing and explaining human behaviour in ways in which knowledge is seen as uniquely personal, and constructed by each person within their social contexts rather than being rule-governed by methods of natural science (Cohen, Manion, & Morrison, 2000). This paradigm therefore fits with a social constructivist epistemology and a critical realist ontology because they all assume multiple interpretations of reality.

An interpretivist approach demands a closeness with the phenomena to allow the researcher to make sense of the context and participants involved in the research.

The thoughts, intentions and actions of participants are understood through their explanations and the articulation of their views. However, the onus is on the researcher to interpret any data via selection, processing, analysis and presentation in a way that is sympathetic to participant views. One of the advantages of working within this paradigm is that it supports close collaboration between the researcher and the research participants, in my case myself as the researcher and my study participants. The development of trust over time meant that participants became very open and honest. This meant that I could give an in-depth and holistic consideration of the rich data that was generated to allow for a better understanding of their thinking and sense-making about their own development.

Adopting a qualitative interpretivist approach does not dictate specific data generation methods. It is possible to draw on several methods enabling the triangulation of data. This study used a range of data sources, each one providing data that contributed to different aspects of the research and participants' social reality (Cohen et al., 2000). As explained by Denzin and Lincoln (2011):

> Qualitative research involves the use and collection of a variety of empirical materials – case study, personal experience, introspection, life story, interview, artefacts, and cultural texts and productions, along with observational, historical, interactional and visual texts – that describe routine and problematic moments and meanings in individuals' lives. Accordingly qualitative researchers deploy a wide range of interconnected interpretive practices, hoping always to get a better understanding of the subject matter at hand. It is understood however that each practice makes the world visible in a different way. Hence there is frequently commitment to using more than one interpretive practice in any one study. (p. 3)

Miles and Huberman (1994) discuss the strengths of qualitative data to highlight their "richness and holism, with strong potential for revealing complexity. Such data allow for 'thick descriptions' that are vivid, nested in a real context, and have a ring of truth that has strong impact on the reader" (p. 10). The aim of this study was to generate a deeper understanding of some beginner teachers' experiences and meaning making through rich and "finely nuanced accounts of human action" (Gergen & Gergen, 2000, p. 1027).

To conclude this section, it is acknowledged that the researcher and participants live in a social world, and are influenced by a range of external and internal factors, including others. An interpretative qualitative approach drawing on social constructivism was adopted as suiting the aims of this study given that it investigated the development of teachers' summative assessment literacy over time within a range of contexts and drew on the teachers' own interpretations of their development as teachers.

4.2   The participants

Qualitative research requires careful purposeful selection of participants to best help the researcher answer the research questions (Creswell, 2012). In this study the selection of participants (referred to as beginner teachers) was purposeful, based on the beginner teachers being typical of those working in New Zealand, my ability to observe their shared learning experiences, and convenience (in that all beginner teachers were based in the same city and the same university programme and the university of which I am part). These aspects allowed me to generate rich information and the choice was convenient. The non-representative nature of purposeful sampling was seen to be the best way to learn about the phenomenon under investigation, on the basis that the data generated would be "information rich" (Patton, 1990, p. 169) with respect to the research questions, and manageable. A disadvantage of this selection was that there was less variation in the ITE input the beginner teachers received given that they were all in the same ITE programme. The selection method is best described as "typical" (Creswell, 2012, p. 208) in that it represented people typical to a secondary science teacher ITE programme in New Zealand i.e., a mix of recent graduate and career-changers, with a 3:5 ratio of men to women.

My access to beginner teachers was with the agreement of administrators and staff working in the ITE programme at a university. Only when this had been granted did I approach potential beginner teachers. The beginner teachers were science graduates enrolled in a one-year Graduate Diploma of Teaching (Secondary) programme. All those enrolled in the compulsory secondary science curriculum course were approached and offered the opportunity to be involved in this research project. They were initially visited in a university science education class

in the first week of lectures, 18 February 2013, and a general explanation of the study was given to the class, with the lecturer's consent. Potential participants were given a letter of invitation (Appendix A) which introduced the proposed research and my background as a researcher, and included a consent form (Appendix B). An information sheet (Appendix C) was provided which explained the research questions, the nature of participant involvement and the methods used to obtain data, the duration of the study, participants' rights within the study and ethical considerations, including confidentiality, potential for harm, conflicts of interest, procedures for handling information and materials produced from the course of this research, complaints, and dispute resolution.

Table 1: Participant Details

Beginner teacher	Age	Male/Female	Qualifications	Previous professional employment
Bella	21	Female	B.Sc	
Benjamin	24	Male	B.Eng	Engineer
Elisa	21	Female	B.Sc	
Kate	23	Female	B.Sc (Tech)	
Mary	26	Female	B.Sc	Personal trainer
Susie	25	Female	B.Sc (Tech)	Scientist
Ryan	52	Male	B.A, B.Sc	Construction worker
Wiremu	24	Male	B.Sc	

The following week another visit was made to the class, where I answered all follow-up questions individually and collected the signed consent forms of all beginner teachers willing to be involved in the research. All eight members of the class agreed to be involved in the study and all were successful in securing a teaching position in 2014, so they agreed to remain in the study until data collection concluded in mid-2014. I felt that eight beginner teachers was sufficient yet manageable, given the amount of data generated through interactions with each person. I did not want to decline those who had volunteered. A summary of the beginner teachers, their ages, qualifications and previous professional experience is given in Table 1 above. Pseudonyms are used for all teachers. Quotes from the interviews reported in the following chapters are labelled with a

teacher identifier and an interview identifier for reference e.g., a quote followed by [Ryan, I5] indicates Ryan provided the quote during Interview 5.

4.3   Data Collection

My aim throughout this study was to document the teachers' summative assessment literacy development. This section describes the methods used to gather data. Firstly, an outline of the timeframe is given. Then methods used in the study are described: interviews, questionnaires, artefact collection and observation.

4.3.1   Time frame

The research was conducted in two phases to gather data at different points of teachers' careers: Phase 1 – Initial teacher education, and Phase 2 – Working as a teacher. Data collection occurred over an 18 month period. Table 2 summarises the research phases and data collection methods used to inform the research.

Table 2: Research Phases and Data Collection Methods Used

	Beginner teacher activity	Data gathering	
Phase 1 – Initial teacher education 2013			
February – April	University courses	Interview 1 Artefact collection 1 Questionnaire 1	Observations over 8 weeks in university classes
April - May	Practicum 1		
June - August	University courses	June (post-practicum 1) Interview 2 Artefact collection 2	Observations over 6 weeks in university classes
		August (pre-practicum 2): Interview 3 Artefact collection 3	
August - September	Practicum 2		
October - November	University courses	Interview 4 Artefact collection 4 Questionnaire 2	Observations over 2 weeks in university classes
Phase 2 – Working as teacher 2014			
February – July	Working in schools	June/July: Interview 5 Artefact collection 5 Questionnaire 3	

4.3.2   Data generation methods

In order to generate data to answer the research questions a number of data gathering methods were used in order to build up a rich description of the experiences and constructed meanings of the beginner teachers involved. Interviews provided the main source of data, while questionnaires, evidence from artefact collection and observation were also used for the purposes of triangulation and to offer insights into the same phenomena from different perspectives.

4.3.2.1   Semi-structured interviews

Interviews are a key way to explore how people experience and understand their world, as interviews allow them to describe and make sense of (theorise) their activities, experiences, ideas and opinions in their own words (Kvale, 2007). In this study interviews allowed for the investigation of issues, ideas and phenomena that were complex and entailed a number of perspectives.

Interviews can be very structured and formal, using response categories to produce data that can be statistically analysed. Structured interviews are useful especially when the interviewer is aware of what she does not know and is in a position to ask questions which will fill the knowledge gaps (Cohen et al., 2000). Alternatively, semi-structured interviews can be used when the researcher is less sure of what she does not know and relies on the interviewees to contribute to the conversation, telling stories and co-constructing meaning with the interviewer. This type of interview uses the framework of a set of core questions for all participants, but the interviewer has the freedom to use follow-up questions that develop or build on what has been discussed. This allows for more in-depth explanation or storying from the participant. It means that all participants have the opportunity to talk about a range of topics that the interviewer is interested in, and that the interviewer has the opportunity to probe for clarification, and to follow up on ideas that might not have been anticipated (Brenner, 2006). Semi-structured interviews also provide the interviewees with the opportunity to use their own words, and to, in some way, direct the conversation to aspects of the research which interest them, including where the interviewees perceive gaps, contradictions and difficulties (Banister, Burman, Parker, Taylor, & Tindall,

1994). The interviewees are able to clarify and elaborate on their ideas in order to provide the interviewer with a richer understanding of their views on the topic under discussion. During semi-structured interviews the ideal outcome is the co-construction of negotiated and shared meanings by both the interviewer and interviewee. The design of interview questions that encourage participants to talk expansively requires care and planning so that data is generated to address the research questions and the research objectives are met (Brenner, 2006). In this study semi-structured interview was selected as the main method of data collection in this study as it afforded flexibility to pursue further issues of interest that were unanticipated. As suggested by Patton (2002), the interviews began with descriptive questions close to the beginner teachers' current experience. The interview schedule used is available in Appendix D.

Interviewing is challenging as it requires participation by both researcher and interviewee, so the question of power relations needs to be considered, especially for open and semi-structured interviews. An important aspect of any interview process is the development of a positive relationship between interviewer and interviewee so that the interview is not merely a data collecting exercise but a meeting of two people who have shared interests and concerns (Brenner, 2006). Interviewees need to feel secure to be able to talk freely. True objectivity on the part of the interviewer is not possible given their personal, cultural and professional repertoires, and this needs to be acknowledged rather than ignored (Qu & Dumay, 2011). Given these concerns it was very important that I was reflexive throughout the research process so that the views of the interviewees were represented as they intended.

In this study, an iterative process was used through the series of five interviews in order to deeply explore the issues over time (refer to Table 2). The spaced interview scheduling not only allowed for the documentation of beginner teachers' development over time, it also allowed for time triangulation for the stability of data. The interviews used a *past-present-future* structure similar to that used by Horn and colleagues in their investigations of novice teachers' motivation to learn and use various teaching strategies (Horn, Nolen, Ward, & Campbell, 2008; Nolen, Horn, et al., 2011; Nolen, Ward, & Horn, 2011) (see Table 3). This meant some questions focused on the past experiences and understandings of the

113

beginner teacher, some on the present, and some on future experiences, understandings, hopes and expectations. In each interview preceding *present* experiences and understandings from earlier interviews were re-examined as *past* experiences, and previous *future* anticipated experiences and understandings were revisited as *present* topics.

Table 3: Interview Structure Used in this Study

Interview	Discussions focusing on:					
Interview 1	past	present	future			
Interview 2		past	present	future		
Interview 3			past	present	future	
Interview 4				past	present	future
Interview 5					past	present	future

All interviews were transcribed by me soon after they were completed. The transcription was analysed to inform the scheduling for the next interview. For example, to allow for the transition from *present* to *past*, quotes of interest were identified from the transcript of a beginner teacher's earlier interview. These quotes were added to the interview protocol and were repeated to the beginner teacher to give them the opportunity to reflect on their past comments. New themes that emerged and were identified as potentially important caused me to generate new questions which were added to the pool of base questions for all beginner teachers. Although the detail of quotes and structure varied slightly for individuals, each interview shared similar protocols. For example, in Bella's second interview I reminded her about what she said about her fear of teaching to the test, and the need to find a balance of telling the students enough, but not too much, seen by her as a midpoint. I asked if she had any further thoughts to add:

> You've talked about that again, not teaching to the test ... [in the last interview] you said *"I don't want to be too vague so that they don't know what's in the assessment, so I guess it is finding the midpoint"*.
> So when you think about that comment, and now after your experiences of working with those year 12s…? [Frances in Bella, I2]

In another example I asked Elisa about her views on science being easier to assess than technology:

> Remember last time we talked about assessment, you were comparing
> science and technology and you were saying that you thought that
> "*science would probably be more clear-cut when it comes to*
> *assessment, than technology*". Did you find that, or do you still think
> that? [Frances, in Elisa, I2]

By using their own quotes from previous interviews, I found I was able to revisit their ideas in a way that was accessible and non-confronting. They all seemed to be relaxed about further explaining their views when I used this technique.

At times, the interviews were quite interactive as I shared my own personal experiences and told stories in order to reciprocate in sharing, as suggested by Phoenix and Sparkes (2009). It was important to be willing to give of myself in acknowledgement of the beginner teachers' willingness to freely share with me. At times, the teachers asked for my personal opinions on educational issues and I did briefly respond usually at the end of the interviews, in order to reduce the likelihood of them replying in kind to the ideas I raised. Trust and openness developed over time, as demonstrated by increased mutual sharing, a process that illustrates the connectedness between researchers and the participants (Lather, 1991). As Lather suggested, I found that over the sequence of in-depth interviews a more relaxed, open relationship developed between myself and the beginner teachers.

4.3.2.2   Questionnaire

Questionnaires are useful instruments for collecting information and are relatively easy and efficient to use in data collection. However, there are limitations to the flexibility and depth of responses elicited as typically questionnaires do not permit elaboration or qualification in the responses given (Cohen et al., 2000; Desimone & Le Floch, 2004; Ho, O'Farrell, Hong, & You, 2006). The trustworthiness of information provided via a questionnaire needs to be considered, as participants may respond in ways they think are most socially acceptable, rather than reflecting their real views (Ho et al., 2006). However, with refinement through piloting a questionnaire it is possible to increase its validity and practicability (Oppenheim, 1992) . Questionnaires can take many forms, but most are able to be administered in a comparatively straightforward manner.

The questionnaire used in this study was the "Beliefs about assessment" questionnaire used in a New Zealand research project: *Learning to become 'assessment capable' teachers 2009–2013* (Smith et al., 2014). This questionnaire was designed to investigate the ideas preservice teachers had about assessment and was used with permission of the authors. The questionnaire was used to triangulate the interview data and alert me and the teacher to views potentially worth following up in the next interview. It was used three times: the beginning and the end of the 10 month ITE programme (February and November 2013) and six months after starting work. The purpose of the initial questionnaire was to obtain information from each beginner teacher from which to generate discussion points in the first interview. The data gathered at the end of the ITE phase was used as a check for changes or shifts in thinking, again to inform interviews. Where considerable shifts were noted, these were further discussed with the associated beginner teacher at the following interview. Similarly, the questionnaire was completed in June/July 2014 to further document teachers' beliefs and understanding of assessment. The teachers reported they found the questionnaire straightforward to complete.

4.3.2.3 Artefact collection

Documentary evidence allows researchers to supplement primary data, and may itself be primary data (Cohen et al., 2011). People represent themselves and their practice in a number of ways including through objects and "the very act of preferring one option to another gives a researcher insight into the basis of the selection" (Newby, 2014, p. 155). Beginner teachers in this study were invited to provide evidence of their summative assessment knowledge and practice by giving me artefacts that they felt illustrated their summative assessment practice. The artefacts they provided included copies of assessments that they had developed and used, lesson plans, marked quizzes, files of NCEA tasks and so on. The artefacts themselves and their analysis provided supplementary stimulus material for discussion during interviews. They helped with triangulation by providing another data source for the study. As discussed by Cohen et al. (2000), documentary evidence may have face validity but may be difficult to interpret and may contain material that is irrelevant; hence, there was real value in discussing these artefacts with beginner teachers who offered them.

4.3.2.4   Observation

Observations are a powerful way to gain insight into a situation (Cohen et al., 2000). Observations can give access to phenomena that are often obscured (e.g., non-verbal clues), or that are not repeatable (e.g., specific discussions or debates within groups) (Banister et al., 1994). As discussed by Angrosina and Rosenberg (2011), observation implies some level of collaborative research with participants, given their ability to discuss what has been observed and hence help the researcher carry out the research plan. Observation in this study enabled me to experience the learning context of the beginner teachers in university ITE courses including the input they received from lecturers and classmates.

As a trial in June 2012 I had attended a number of classes in a science curriculum course with a previous cohort of beginner teachers, and this encouraged me to pursue this method within the study. The class discussion in the pilot raised several issues that I had not considered as important in the development of beginner teachers' summative assessment literacy, including assumptions I had made about their prior knowledge. For example, the beginner teachers appeared quite confused over the NCEA system. Over the course of the study I attended multiple ITE classes as a non-participant observer in order to experience some of the input and activities that beginner teachers were involved with during their ITE. I attended all classes where the lecturers indicated to me that assessment would be a focus of the session. This resulted in my gaining a deeper understanding of some of the input and experiences that the beginner teachers had during their ITE. I was aware that my presence in classes could change the dynamics of the situation being observed, so I made sure to sit to the side and near the back of the classroom, and to keep my interactions with beginner teachers and lecturers to a minimum.

The observations were unstructured. During these observations I paid attention to times that lecturers and beginner teachers talked about assessment or assessment related matters. Notes were recorded on paper for the purpose of summarising ideas that emerged in the classroom. These observations proved helpful during the interviews as the beginner teachers knew I was familiar with what they were referring to.

117

4.4   Data analysis

It is through data analysis that sense is made of the data, and through subsequent representation by the researcher that participants' stories are told (Merriam, 1988). For qualitative studies decisions around units of analysis have to be made. Often in qualitative research, multiple units of analysis fit within a nested arrangement, that is, there is a broader level within which a number of instances at a narrower level are positioned (Cresswell, 2012; Patton, 2015; Yin, 2016). In this study there are two levels of analysis: the broader level is the cohort of beginner teachers engaging in ITE in a particular setting over an 18 month period, and the narrower unit of analysis is at the level of individual teacher. Following analysis, the representation of data requires researchers to make many decisions about what to include by way of description, quotes and excerpts from participants. This means the representation of participants' thoughts and ideas is a somewhat limited version of what was experienced and found, nonetheless the data is still valid (Eisenhart, 2006; Kvale, 2007; Miles & Huberman, 1994).

In this study data analysis was carried out during and after data generation and focused on the making of meaning through multiple readings of the transcripts of interviews, questionnaires, field notes and documentary evidence in the form of artefacts. An iterative process of inductive and deductive reasoning was used to categorise and reduce data and develop codes, as is discussed in the following section.

I fully transcribed all interviews soon after they occurred while they were fresh in my mind. All recordings were clear so the transcription process was straightforward, although it is acknowledged that I made decisions regarding punctuation. The transcriptions were generally verbatim except when the content of the conversation was deemed to be irrelevant to the study. Notes regarding other features of the interviews were added where it was deemed needed to more authentically communicate meaning; for example, laughter, and pauses while thinking. Within a week of each interview, the transcript was returned to each beginner teacher for checking. Beginner teachers were invited to add or change content in the transcripts to ensure that it fairly reflected the meanings they wanted to convey. No changes were made or requested by the teachers.

Table 4: Data Analysis Process Timeline

Data analysis timeline	Interview	Analysis (in preparation for next interview)*	Condense codes, review over entire data	Rubric development	Rubric refinement
March 2013	Interview 1				
April		Analysis			
June	Interview 2				
July		Analysis			
August	Interview 3				
Sept		Analysis			
Oct–Nov	Interview 4				
Nov-March		Analysis	Analysis	Analysis	
July/Aug 2014	Interview 5				
Aug–Oct		Analysis	Analysis		Analysis
Jan–March 2015					Analysis

*Note: As new codes were developed through the analysis process, earlier data sets were re-coded to include these.

Interviews were the primary source for data analysis. Overall, the transcriptions of 40 recorded interviews were analysed. I used the computer software NVivo (Version 10) to help manage and code data according to themes.

The analysis occurred in two stages. The first stage involved the analysis of data from each interview in preparation for the next interview in the sequence, and included the development and use of categories and codes (see Section 4.4.1). The second stage involved the development and application of a rubric to trace teacher development over time (Section 4.4.2). The analysis also involved a continuous process of drawing and verifying conclusions. A summary of the data analysis process is given in Table 4.

4.4.1　Content analysis

Content analysis has been broadly defined by Patton (2002) as "any qualitative data reduction and sense making effort that takes a volume of qualitative material and attempts to identify core consistencies and meanings" (p. 453). This analysis process is systematic, allowing for both inductive and deductive approaches. Its flexibility means that analysis does not necessarily proceed in a linear fashion, as its concern is in discerning meanings, intentions and consequences within context (Elo & Kyngas, 2008). Meanings emerge through subsequent extensive engagement with the data over time. Categorisation, coding and reduction are features of the analysis and allow meanings to be sorted and organised so they can be reported (Miles, Huberman, & Saldana, 2014). Codes are labels developed for units of meaning within the data, and in this study were used to help manage the analysis process. Affixing codes to data enables categorisation and the identification of emergent themes (Miles & Huberman, 1994).

Content analysis commenced as the first stage of analysis. I immersed myself in the data initially by attending to individual transcripts after each interview round, summarising tentative key ideas and identifying themes. As recommended by Miles and Huberman (1994), I started the process of creating codes by using a provisional "start list" (p. 58) which was formulated by using the model for science teachers' assessment literacy proposed by Abell and Siegel (2011). The list comprised: knowledge of purposes of assessment, knowledge of what to assess, knowledge of strategies and design, knowledge of assessment interpretation, principles of assessment, learning about assessment. This initial deductive coding was followed by and merged with subsequent inductive coding as new codes emerged progressively during data analysis.

Because of the iterative nature of the interview design I had to work on data analysis on the completion of each interview in order to prepare for the following interview. Additionally, in preparation for the following interview, particular quotes were selected for which I wanted further comment from the beginner teachers. This analysis process was repeated after each of the interviews. The spaced gathering of data gave me the opportunity to observe the development of teachers' thoughts and ideas over time, and my interpretations of data collected later in the study caused me to re-visit and revise codes or re-code earlier data

sets. When any new code was developed during the analysis I re-read all the previously coded data to re-code for the new code. For example, the code "Preparing students for standards-based assessment" was added at Interview 4, and then all previous interview data was recoded.

After the analysis of the fourth interview, at the end of the first year of data gathering, the data was reduced through a process of "selecting, focusing, simplifying, abstracting and transforming" (Miles & Huberman, 1994, p. 10). All coded data was read, and where overlaps or common content was identified codes were conflated, and in some cases the descriptors used to identify them were changed to better represent the content. For example, the three codes "benefits or effects of assessment for students", "benefits or effects of assessment for teachers", and "students can monitor their progress" were combined in the code "understanding assessment consequences". The use of NVivo facilitated coded data being retrieved and grouped over interviews. This process consolidated the number of categories as clusters of themes which were grouped under headings that summarised their focus (see Appendix E). This was then reviewed after the final data set was collected and all data was checked and final adjustments made of all coded data. In total 12 condensed codes resulted, grouped into four categories listed below (see Appendix F for sample quotes for each category).

- Knowledge of assessment
  - Ability to define assessment
  - Knowledge of purposes of summative assessment
  - Knowledge of what to assess
  - Knowledge of strategies and design
  - Knowledge of assessment interpretation
- Understanding the context for assessment
  - Preparing students for standards-based assessment
  - Using summative assessment formatively
- Recognising the impact of assessment
  - Understanding assessment consequences
  - Fairness
- Self-analysis of personal development of assessment literacy
  - Enablers and barriers
  - Personal responses as a teacher

One challenge I faced in the coding phase was discerning between what beginner teachers knew (teacher knowledge) and what they believed (teacher beliefs). As discussed in the literature review, the development of teachers' professional knowledge is a foundational aspect of becoming a teacher. However, it was very difficult to separate knowledge and beliefs in the data. Because of the tacit nature of knowledge, it is difficult to explicate, as people may not be aware that they have access to knowledge, or are developing knowledge. For the purposes of coding, the following definition of teacher knowledge was used: "All that a person knows or believes to be true, whether or not it is verified as true in some sort of objective or external way" (Alexander, Schallert, & Hare, 1991, p. 317).

To illustrate the coding, here is a sample of text with codes:

> So there's diagnostic, formative, and summative (*Knowledge of purposes of summative assessment*). Assessment doesn't have to be just tests and the internals and the externals, but you can have a lot of informal ones (*Knowledge of strategies and design*)… It can be fun, it doesn't have to be boring, and scary and stressful. You can make them exciting, not really the externals but you can make the internals more exciting as you link them to context so it can be more exciting. Assessments are good, so you can gauge where the students are at and you know where to help them and see where the students need attention (*Using summative assessment formatively*). [Bella, I4]

As a result of the time spent immersed in the data, and because I conducted all interviews and completed the transcriptions myself, I knew the data very well and was very aware of the wider conversations and contexts around which the data were generated. This gave me considerable confidence that the representation of the beginner teachers' ideas into 12 categories was trustworthy. A sample of my coding was checked and confirmed by two colleagues to ensure consistency and credibility in my work. At the end of the data gathering period I reviewed all coding, checking for consistency over time. Through this process it was important to critically reflect on myself as researcher, the "human as instrument" (Guba & Lincoln, 1981), and I practised reflexivity by regularly reflecting on my own position as a researcher and the way I engaged with all aspects of the data analysis.

4.4.2 Generating a rubric to trace teacher development over time

The second stage of data analysis involved the development and use of a rubric to track assessment literacy development over time. This rubric was an important element of the study as its use permitted beginner teacher development to be documented and presented visually. The rubric was developed using a combination of deductive observations from the literature and inductive observations from this study.

4.4.2.1 Deductive development

I consulted the literature to develop the initial Summative Assessment Literacy Rubric (SALRubric) by (i) identifying relevant dimensions of skills and knowledge of summative assessment literacy, (ii) articulating levels of expertise from novice to expert, (iii) developing a framework of descriptors, and (iv) considering the New Zealand context in which the SALRubric would be used. Each step is discussed in turn.

Abell and Siegel's (2011) developmental model for science teachers' summative assessment literacy (MSTAL) was based on conceptions of PCK and the assessment triangle (Pellegrino et al., 2001). Incorporating findings from their own work, the MSTAL has, at its core, teachers' assessment values and principles that are grounded in teachers' views of learning. Abell and Siegel argue that these aspects influence teachers' use of knowledge in decision making. Their model provided an initial framework for the SALRubric across multiple dimensions comprised of knowledge of four aspects of assessment, values in assessment and teacher orientation (Abell & Siegel, 2011; Lyon, 2013a, 2013b; Schneider & Plasman, 2011; Siegel, 2014).

Dreyfus and Dreyfus' (1986) model described levels of expertise existing in skill acquisition development that mapped routines and decision making from novice to expert levels. As discussed earlier, a key element of teacher development illustrated in this model was the shift in teachers' thinking by adhering to rules or 'knowing that', to being able to flexibly adapt or apply principles in practice or 'knowing how' (Dreyfus, 2004; Dreyfus & Dreyfus 1986; Eyers, 2014; Schneider & Plasman, 2011). Lyon (2013a) developed a framework and set of descriptors for developing assessment expertise in pre-service teachers working in linguistically diverse classrooms. His work informed the development of

123

descriptors for the SALRubric, through the idea of grouping codes under headings which signify important analytical lenses.

Finally, elements of policies and practices related to summative assessment in the New Zealand context, in particular the requirements with respect to secondary school qualifications (NCEA) and the strong policy focus on formative assessment, were considered to ensure the SALRubric was appropriate to teachers working in New Zealand.

4.4.2.2 Inductive development

An iterative process of refinement has been found to be useful in constructing learning progressions (Furtak, Morrison, & Kroog, 2014) and rubrics (Lyon, 2013a). Draft dimensions for the SALRubric were reconsidered in response to the analysis of the data generated by the beginner teachers, particularly after later data collection points. This allowed other important dimensions to be added and descriptors to be more finely tuned. Decisions about which dimensions would be included were influenced by their importance in the assessment literature, their importance in the New Zealand context, and their relevance and importance for teachers as evidenced in data gathered from beginner teachers. As a result of this refinement, dimensions that focused on using summative assessment for formative purposes and on assessment consequences and fairness were added to earlier versions of the rubric. Further refinements were made to the wording of criteria in the SALRubric based on expert advice from within New Zealand and overseas when the work was presented, and from the Practising Teacher Criteria (Education Council, 2015). Initially three levels of competence were identified for the dimensions of the SALRubric from the data: novice, advanced beginner, competent. However, after the final interviews and after close reading of documents from the New Zealand Ministry of Education and *Practising Teacher Criteria* (Education Council, 2015), two further levels (proficient and expert) were added to better represent summative assessment literacy for all teachers in New Zealand, rather than just for beginner teachers.

The final version of the SALRubric includes the following 10 dimensions grouped into three categories: Knowledge of assessment, Understanding the context for assessment, and Recognising the impact of assessment.

- Knowledge of assessment
  - Ability to define assessment
  - Knowledge of purposes of summative assessment
  - Knowledge of what to assess
  - Knowledge of strategies and design
  - Knowledge of assessment interpretation
- Understanding the context for assessment
  - NCEA assessment
  - Preparing students for standards-based assessment
  - Using summative assessment formatively
- Recognising the impact of assessment
  - Understanding assessment consequences
  - Fairness

A fourth category resulting from condensed codes involved beginner teachers' self-analysis of their personal development of assessment literacy (Section 4.4.1), but this was not included in the SALRubric as this category was more concerned with self-perception through reflection than with development of summative assessment literacy per se. The reflection evident in the beginner teachers' comments was a contributor to their development of summative assessment literacy rather than an aspect of their assessment literacy per se. The self-analysis category is instead reported on in Chapter 8, where the focus is on what beginner teachers perceived as contributors to their learning.

Each dimension described five levels of competency: 'novice', 'advanced beginner', 'competent', 'proficient' and 'expert'. These were anchored at the novice level with what was observed as the lowest level of summative assessment literacy on beginner teachers' entry to ITE, combined with the pattern seen in other rubrics (e.g., those developed by Lasater, 2007; Lyon, 2013b). The expert level was anchored by the New Zealand Education Council's criteria for practising teachers (Education Council, 2015), and the expectations of teachers reaching Advanced Classroom Expertise Teacher Recognition (TeachNZ, 2006). Table 5 summarises the broad definitions of the five levels of competence from novice to expert. The SALRubric is detailed in Table 6.

Table 5: Levels of Competence Used in SALRubric

Level	Definition
Novice	Limited competence in dimensions of summative assessment literacy, where a teacher either does not consider the relevance of the dimension, or shows limited understanding, and follows basic assessment tasks provided by others.
Advanced beginner	A general understanding of most of the dimension at a basic level, taking notice of the context but still detached from the learning situation.
Competent	An overall understanding or competency in the dimension, becoming more emotionally invested in the context, but yet to demonstrate the nuanced detail inherent in the dimension.
Proficient	A clear understanding of the complexities of the dimension within the contextual framework, approaching decision making from the perspective of deciding between options, applying analytic reasoning.
Expert	A thorough understanding of the detail of the dimension within its context, applying intuitive decision making, and having the capability to contribute understandings to the wider teaching community.

Table 6: Summative Assessment Literacy Rubric (SALRubric) Indicating Levels for Dimensions of Summative Assessment Literacy

Novice	Advanced beginner	Competent	Proficient	Expert
Knowledge of assessment				
1. Ability to define assessment				
Can describe assessment in very broad terms based on own experience.	Can define summative assessment, formative assessment, diagnostic assessment and their purposes using simple descriptions.	Can define assessment purposes and their effects and uses in a more integrated way, understanding that teachers' ethical decision making is required.	Can define assessment and considers its forms and functions and its effects within the class. Aware of complexities involved in assessment decision making including fairness and ethics, and issues of authenticity.	Can define assessment and demonstrates critical engagement with assessment dimensions, focusing on awareness, impact, and challenging the naturalness of assessment practices within assessment community.
2. Knowledge of purposes of summative assessment				
Considers summative assessment purposes in vague terms, without using any specific terminology or without identifying specific purposes.	Considers specific purpose/s for summative assessment such as accountability, qualifications, schools uses, informing pedagogical decisions.	Considers a range of specific purposes for summative assessment including reporting progress, informing pedagogical decisions, qualifications, school uses. Understands that the same assessment task can serve a range of purposes.	Considers accountability, reporting, qualifications and informing pedagogical decisions as valuable purposes of summative assessment as it links to teaching and to the New Zealand Curriculum (NZC).	Considers a wide range of specific purposes for summative assessment including reporting progress, informing pedagogical decisions, qualifications, school wide/national uses. Makes contributions of the creative use of summative assessment for a range of purposes to the wider teaching community.

Novice	Advanced beginner	Competent	Proficient	Expert
3. Knowledge of what to assess				
Considers assessment as a generic activity and does not yet link assessment to the specifics of what has been taught.	Considers that some science knowledge is important to assess or that what is taught needs to be assessed. Understands that NZC provides levels that must be assessed against, and that NCEA standards provide specific assessment criteria.	Considers specific science knowledge and skills that are important to assess, and is comfortable in critiquing assessments for this content. Links assessment to what has been taught. Considers the importance of linking assessment to standards criteria.	Considers the need to identify specific science skills and knowledge that are important to assess. Can see need for alignment between NZC, teaching and assessment. Considers validity. Can accurately critique for content within assessment tasks, including NCEA tasks.	Considers their use of an in-depth knowledge of science content and NZC, and learning progressions when assessing science. Can identify critical content to assess for a range of topics, and can critique assessment tasks for this content. Considers their contribution of this knowledge to the wider teaching community.
4. Knowledge of assessment strategies and task design				
Describes generic forms of assessment.	Considers assessment types and uses prepared materials to assess students. Considers the need to use own ideas and a range of task types but is not yet confident designing own tasks.	Considers the need for assessment tasks to match specific learning objectives (SLOs), and considers the need to adjust/adapt materials to suit context and students. Considers reliability and validity in task construction.	Considers the design of assessment tasks that best match SLOs/ standards and students that use authentic scientific practices. Understands the need for assessment tasks to be critiqued for validity and reliability.	Considers the design of assessment tasks that best match SLOs/ standards and the need to create a range of assessment tasks that engage students in authentic scientific practices e.g., open inquiry. Understands the need for assessment tasks to be critiqued for validity and reliability. Willing to share these with the teaching community.
5. Knowledge of assessment interpretation				
Considers the use of a marking scheme directly to give a grade.	Considers importance of using marking scheme accurately and making judgements based on this.	Considers importance of using marking scheme accurately and making judgements based on interpretation of evidence.	Considers importance of using marking scheme accurately and making judgements based on evidence. Considers individual students' contexts when interpreting assessment data.	Considers clear accurate judgements about students based on summative assessment evidence, using marking scheme. Thinks critically about what assessment results mean –in terms of child, school, community.

Novice	Advanced beginner	Competent	Proficient	Expert
Understanding the context for assessment				
6. NCEA assessment				
Understands basics of NCEA without knowledge of particular standards, or teachers' specific roles in NCEA assessment.	Considers and uses pre-written assessments tasks for NCEA internal standards but not comfortable in writing own tasks. Has some confidence in marking and moderation, but still needs guidance.	Considers the need for alignment between teaching, NCEA standards and assessment tasks, and understands the marking and moderation of NCEA tasks. Adjusts/adapts assessment tasks to suit class/students.	Considers all aspects of NCEA assessment with confidence, with ability to set and critique NCEA assessment materials that meet standards, and consider and adapt tasks to match students' needs. Understands and uses moderation processes.	Understands and can critically engage with all aspects of the NCEA system and can communicate recognised best practice in NCEA with teachers in wider teaching community.
7. Preparing students for standards-based assessment				
Does not consider the teachers' role in developing students' assessment literacy or preparing students for assessment tasks.	Considers the needs students have to understand assessment processes and procedures. Considers sharing assessment criteria with students.	Considers the teacher's role in preparing students for assessment. Considers sharing assessment criteria with students, using practice tasks, marking these and providing feedback.	Considers the teachers' role in preparing students for assessments including developing formative/ practice tasks and providing feedback and feed forward to students on their performance. Helps students develop deeper understanding of successful engagement with assessment.	Considers it part of the teachers' role to prepare students for assessment tasks by ensuring they understand the expectations around criteria and marking/ feedback, use of exemplars. Considers the development of student assessment literacy. Contributing this practice to the wider teaching community.
8. Using summative assessment formatively				
Taking action as a result of summative assessment is not considered.	Considers teachers can learn from summative assessment including NCEA in general terms, and that action may result.	Considers summative assessment evidence including NCEA as an important source of information for teachers, for feedback on their teaching, for feed forward for further teaching, planning.	Considers summative assessment results including NCEA and the appropriate analysis of these as important data on which to act with their class and with individuals.	Considers all summative assessment to be useful data from which to learn and can analyse individual and class data. Can articulate findings and actions from their analysis of summative assessment data and considers using this for future planning in class, school, community.

Novice	Advanced beginner	Competent	Proficient	Expert
Recognising the impact of assessment				
9. Understanding assessment consequences				
Does not consider that assessment really affects people or has consequences for them.	Considers the effects of assessment on students (e.g., emotional, motivational), or that assessment has effects for teachers.	Considers effects of assessment on students and uses them to students' advantage /mitigates harm, and considers the effects of summative assessment results on teachers/schools.	Considers the effects of assessment on students and teachers – uses these to mitigate negative effects and to benefit student learning.	Considers the consequences of assessment on teachers and students, and uses these for students' benefit. Considers wider impacts of assessment and assessment policy and communicates these to wider community.
10. Fairness				
Considers that fairness means treating all students identically.	Considers fairness as an important issue in assessment (or the assessment system), but does not feel able to adapt assessment materials for individuals. Considers it is more important that all students are treated the same or did not mention need for differentiation in assessment.	Considers fairness when assessing students. Considers the need to identify ways that would enable students to provide evidence of their learning. Willing to investigate different modes to assess students against the same SLOs/standards, but with a limited range of skills	Considers the complexity of fairness issues in assessment and the need to design /adapt assessments to allow access for all students. Considers "giving opportunity to meet the criteria" more important than having the same tasks (for students). Considers the adaptation of assessment tasks to give students access/equity.	Explicitly considers fairness and access for all students. Tailors assessment in order to give all students the best opportunity to demonstrate their learning. Contributes these understanding to others in the teaching community.

4.4.2.3   Quality and the SALRubric

Moskal and Leydens (2000) raise the challenges of ensuring validity and reliability in scoring rubric development, and these were considered in the development and use of the SALRubric. With respect to validation of the SALRubric, it was important to consider the accumulated evidence that supported the descriptions of beginner teachers' summative assessment literacy. The inductive element of the SALRubric's development used assessment-related evidence including questionnaire and interview data, documented summative assessment planning linked to unit planning, assessment tasks, marking schedules, judgement statements produced by the beginner teachers and marked student work, alongside discussions with the beginner teachers about the artefacts. This array of content-related evidence reflected the beginner teachers' knowledge of the content area being assessed. Care was taken to ensure that the evaluation criteria within the rubric measured factors related to the construct of interest, that is, summative assessment literacy, establishing construct validity. The adjustment of criteria made later in the development of the rubric meant better alignment between the descriptors and the construct of summative assessment literacy. All of the evidence used when developing the SALRubric's dimensions and descriptors was generated while the beginner teachers were in an ITE programme or while they were working in schools, so it is likely that the SALRubric scores reflect practices in the field, that is, high scores using the SALRubric should suggest high scores when beginner teachers are working in their own classrooms.

Reliability of the SALRubric was also a focus in its development. Well defined and clear descriptors assist with consistency in a rubric's use. In the case of the SALRubric, a number of colleagues and assessment experts were asked to review the level descriptors for dimensions, and their advice aided the clarification of the descriptors. Because of the size of the study, I alone used the SALRubric to evaluate teachers' levels of assessment literacy. In order to ensure reliability in judgements, two independent senior academics with expertise in assessment from a university were asked to independently score a sample of data using this SALRubric. Their judgements aligned with mine in almost all cases and provided confirmation of the judgements that I made. I discussed the few discrepancies that did arise in order to confirm where the judgements should lie.

4.4.2.4 Use of the SALRubric

The SALRubric was employed to document and visually map changes in each beginner teacher's summative assessment literacy. For each data gathering time point, the data were carefully analysed and considered as a whole for each beginner teacher, and a judgement of their level of competency was made for each of the 10 dimensions. The combination of interview transcripts, artefacts and questionnaire responses allowed triangulation and enabled me to make holistic judgements regarding which stage in the rubric best matched each beginner teacher's summative assessment literacy development for each dimension at each time point. Teachers were scored 1–5 for each of five data collection time points with 1 being novice and 5 being expert. Where no evidence was available a score of 0 was recorded. Because of the iterative process involved in the SALRubric development, after the final refinement all data was reconsidered and the extra dimensions were scored.

The SALRubric was particularly useful, firstly, in allowing individual beginner teacher development to be tracked and reported in Chapter 7 so that a holistic picture of each teacher's development over time could be summarised visually and secondly, in enabling patterns in the development of each dimension to be considered across the group of beginner teachers. The rubric scores were illustrated for individuals using radar charts, enabling patterns to be more easily recognised.

4.5 Ensuring the quality of the research

The interpretive researcher working in a qualitative frame generates rich data and then interprets the data, as "the primary instrument in qualitative case study research is human, all observations and analyses are filtered through one's worldview, one's values, one's perspectives" (Merriam, 1998, p.39). In qualitative research quality and rigour is demonstrated through trustworthiness (Guba & Lincoln, 2005). In seeking trustworthiness researchers must attend to four criteria: credibility, transferability, dependability and confirmability (Lincoln & Guba, 1985).

4.5.1 Credibility

Credibility involves establishing that the results of a study are believable. It deals with the accuracy of identifying and describing the subject of the study. To

address credibility it is important that the research is planned carefully so that readers can see how well the data and processes of analysis address the intended focus. In this study a number of strategies were used to ensure the study is credible: triangulation, prolonged engagement, member checks, and peer validation.

Triangulation involves the use of a number of methods when generating data about a phenomenon, as a way of adding "rigour, breadth, complexity, richness and depth to any inquiry" (Denzin and Lincoln, 2013, p. 10). The use of multiple methods can help confirm emerging findings (Sarantakos, 1998). Although interviews were the main data source in this study, a number of other sources were used to generate further data and corroborate findings. In particular, questionnaires and artefacts added to a richer data set, and confirmed ideas arising from the interviews. The variety of methods were intended to increase authenticity as Coll and Chapman (2000) note, "by comparing and contrasting one source of information with another the researcher is able to produce a more comprehensive and balanced study in response to the research question" (p. 5).

The data collection took place over an extended period of time (approximately 18 months) so this prolonged engagement helped me to check for stability. As Merriam (1998) indicates, generating data over time, or long-term observation, is a way to increase the credibility of findings. Messages from beginner teachers that were recurrent reinforced their trustworthiness.

In this study I asked the beginner teachers to check the transcriptions of their interviews for accuracy. My tentative interpretations of what they meant, particularly during following interviews, were also shared with the teachers as a further way of allowing them to check for meaning I was taking from the data. Having findings verified by the participants adds to their credibility (Lincoln & Guba, 1985; Merriam, 1988).

Peer validation was also used to ensure credibility. Regular discussion about the research with peers and supervisors was a feature. In particular, regular meetings with my supervisors enabled me to receive critical comment and challenge to my interpretations. Initial findings were presented and well received at a number of

international conferences, at which I invited comments and questions to further probe the study.

4.5.2 Transferability

Qualitative studies are designed to give rich, in-depth accounts of whatever is under investigation and the results are detailed and context specific. This means findings from a qualitative study cannot necessarily be generalised to apply to a whole population, as the selection of participants is not large enough and is not designed to be representative of the population. Instead the concept of transferability relates to how findings may or may not be transferable to another context. In the case of qualitative studies, the author can give suggestions about transferability but it is up to the reader to decide whether the findings are transferable. As Merrian (1998) summarised, "The general lies in the particular; that is, what we learn in a particular situation we can transfer or generalise to similar situations subsequently encountered" (p.211). Patton (1990) describes transferability as something to provide "perspective rather than a truth, empirical assessment of local decision makers' theories of action rather than generation and verification of universal theories, and context bound extrapolations rather than generalisations" (p. 491). In this study, to aid transferability, it was important to have a rich description of the beginner teachers, the context, and the process of analysis together with appropriate quotations. A rich, thick description helps the readers decide what is relevant to them and what findings could be transferred to other situations (Bryman, 2001; Geertz, 1973; Lincoln & Guba, 1985).

4.5.3 Dependability

Another aspect of trustworthiness is dependability, which "seeks means for taking into account both factors of instability and factors of phenomenal or induced changes" (Lincoln & Guba, 1985, p. 299). The context specific nature of this kind of study means that the dependability or consistency of the results obtained from the data should be the focus (Lincoln & Guba, 1985). In this study, prolonged engagement meant there was a particular risk of inconsistency, so I carefully reviewed the way data was treated for consistency over time and kept detailed records of all aspects of the process of this study including decisions relating to question formulation for each beginner teacher, interview scripts, documents and data analysis processes. Keeping an audit trail, which entails detailed record

keeping of all phases of the research project, has been recommended by Bryman (2001) as a way of increasing dependability. I kept the words of the teachers in the forefront of my mind, when formulating codes and themes. This included the reading and re-reading of earlier scripts and notes when working on the data gathered later in the process.

Auditing was also carried out by the university appointed supervisors of this research. This included the requirement for a documented research proposal and ethics application, regular meetings with supervisors, six-monthly progress reports and feedback on all aspects of the research. Bryman (2001) referred to peer auditing, and this was evident in my study by the regular discussions I had with other academic staff involved in teaching about assessment in ITE programmes at the university where I was based.

4.5.4 Confirmability

Confirmability relates to whether findings can be confirmed from the data that has been generated. It is important that dubious evidence or researcher bias has not unduly influenced the findings (Yin, 1988). For this my supervisors were important in that they questioned aspects of my work, monitored my progress and provided me with the opportunity to step back to view the whole work. This included their observations of my use of triangulation to confirm patterns and checking how decisions were made regarding the data that was used. Additionally, my reflexivity, that is, my own regular reflection on my position as a researcher, the practice of the research and the generation of knowledge helped me as I worked on the data analysis and presentation of the findings. As Alvesson (2003) explains,

> Reflexivity operates with a framework that stimulates an interplay between producing interpretations and challenging them. It includes opening up the phenomena through exploring more than one set of meanings and acknowledging ambiguity in the phenomena and the line(s) of inquiry favored, and it means bridging the gap between epistemological concerns and method. (p. 14)

To be critically reflexive meant that I had to examine the assumptions underlying my actions, making my decision making within the study open for critique as well as bending back to look more closely at my own interactions and perspectives from other perspectives (Lincoln, Lynham, & Guba, 2011). It was my intention

that the voices of the beginner teachers be articulated in the findings and to this end I have endeavoured to report these accurately and clearly. I acknowledge, however, that research is a complex activity and that my own views as the researcher will have had some degree of influence in the research process.

Confirmability also concerns the idea of the researcher acting in good faith (Bryman, 2001). It was important that rich descriptions of the experiences of beginner teachers were communicated and that the participation of the researcher 'being there' was acknowledged within these descriptions (Eisenhart, 2006). The ongoing contact between the beginner teachers and myself as the meanings of the research were established also added to this good faith.

4.6   Ethical considerations

Ethical considerations are always of utmost importance when conducting research. Concerns for the welfare and protection of participants need to be foremost in a researcher's mind when planning, conducting and reporting on research (Banister et al., 1994). Social researchers must show respect and take care, with full recognition that aspects of their participants' lives are being observed and written about, exposing them to others (Longworth & McBride, 1994). Ethical issues can arise from the nature of the project itself, the context in which the research is carried out, the methods of data collection, type of data collected, nature of participants and what is done with the data (Cohen et al., 2000). As I embarked on this project I took considerable time in the planning phase to ensure that those involved in my study were informed, would face minimal risk and would be treated with respect.

All arrangements for the conduct of the research were consistent with the University of Waikato's *Ethical Conduct in Human Research and Related Activities Regulations* (2008) including participants' rights to anonymity and their rights to withdraw from the research project. The letter of ethical approval from the Faculty of Education Research Ethics Committee is appended (see Appendix G). Procedures were made available to beginner teachers to allow for resolution of disputes, but these were not called upon.

4.6.1 Informed consent

A key element of ethical treatment in this study was that all beginner teachers involved gave their informed consent to participate. This meant they agreed to be involved, acknowledged that they had a clear understanding of all aspects of the study, and any potential harm from their involvement. They understood that my role was solely as a researcher and that, although employed by the university, I was not involved in any way in the ITE programme in which they were enrolled. They were free to choose whether to volunteer to be involved in the project. There were no repercussions for choosing not to be involved. In this project the key participants were the beginner teachers, but other people could also have been potentially affected; for example, lecturers of university courses, and so care was taken to ensure that these others all freely gave their consent, were protected from potential harm and were fully informed. In this study the beginner teachers were made aware of their right to withdraw from the study at any time, and without question, up until 1 December 2013 for Phase 1 participants, and until 1 July 2014 for Phase 2 participants.

4.6.2 Harm to those involved in the research

All research has the potential to provide benefits or harm to participants. A number of measures were put in place in order to minimise the chance of harm. The Information Sheet (see Appendix C) alerted beginner teachers to potential harm and this was discussed with each of them in the initial meeting.

In order to mitigate potential harm, interviews were only held at times suggested by the beginner teachers. Interviews were not rushed, but neither did they take longer than required, to ensure teachers' time was not being wasted and that they were not stressed. During the interviews the beginner teachers did, at times, disclose information that could potentially harm themselves or others. I had a responsibility to draw beginner teachers' attention to such comments and not use or edit material accordingly. When references were inadvertently made to other people, the identity of those mentioned remained confidential. Beginner teachers were provided with complete copies of their interview transcripts by email. They had the opportunity to check these for accuracy, amend them and make comments before releasing this data for my use in writing this thesis (a form of member checking).

The right of privacy is fundamental, and my participants needed to have control over the information I held about them (Longworth & McBride, 1994). I made sure the information was only used as intended and was stored securely. The beginner teachers have the right of anonymity and confidentiality and care has been taken to ensure the information given does not lead to the discovery of the person, as advocated by Cohen et al. (2000). Similarly, I have not named any lecturers that I observed or who were mentioned in interviews by the teachers. The teachers were not made aware of the others involved in this study, but they may have shared this information about themselves with others.

4.6.3   Reciprocity

In research the interests of participants need to be considered as well as the interests of the researcher (Graue & Walsh, 1998). Social research should result in valuable and useful knowledge and improvement for people, so ethical research must benefit those involved in the research. In this study I, as the researcher, benefited through the completion of my PhD, and it is hoped that future candidates entering ITE programmes may benefit because of the application of the findings in teacher education programmes. The beginner teachers involved in this study also benefited through the opportunities they had to discuss and reflect on their growth, particularly in respect to their development of summative assessment literacy. A number of the beginner teachers acknowledged that they had benefited from our ongoing dialogue. They also talked about how their involvement in this study had helped them be more intentional in their practice.

4.7   Summary

This research adopted a qualitative interpretivist methodology to investigate teachers' developing summative assessment literacy. The study was interpretative in that I sought to interpret and understand how the beginner teachers were developing summative assessment literacy. This study was developed from a social constructivist perspective to acknowledge the way the development of knowledge relies on the social, historical and physical context in which it occurs.

This chapter presented the methodology, methods and analysis employed in this research. Issues regarding ensuring the research's quality and ethical guidelines were discussed. The following three chapters present the findings of the research.

Chapter 5: Summative assessment literacy: Teacher professional knowledge

In the following three chapters the findings of this study are presented. Assessment literacy, as defined by Stiggins (1995) and Willis et al. (2013), involves teachers drawing on professional knowledges in the context of their classroom to make assessment decisions and take action. The literature signalled that summative assessment decisions and action are dependent on teacher assessment knowledge and influenced by a range of factors, termed here as amplifiers and filters (Gess-Newsome, 2015). The concept and action of amplifiers and filters resonates with the core of Abell and Siegel's (2011) model of science teacher assessment literacy: views of learning and assessment values and principles. The following chapters describe the findings relating to beginner teacher perceptions of their professional knowledge, the development of summative assessment literacy for beginner teachers over time, and the amplifiers and filters that influence assessment decision making and practice.

Chapter 5 presents evidence of the breadth of professional knowledges of the beginner teachers in this study. They were able to articulate an awareness of a broad range of dimensions of summative assessment knowledges including knowledge of (i) summative assessment purposes, (ii) what to assess, (iii) summative assessment strategies and task design, (iv) summative assessment interpretation and (v) quality summative assessment. They also demonstrated an understanding of the potential effects of summative assessment on teachers and students.

In Chapter 6 findings are presented that illustrate the development of specific dimensions of summative assessment literacy over time. Development was tracked through the use of the SALRubric (Section 4.4.2). The SALRubric allowed the trajectories of individual beginner teachers to be presented graphically. These trajectories reflect the way individuals experienced and responded to their ITE and first schools. Documented examples of three beginner teachers are presented in more detail to illustrate the variation evident in the development of summative assessment literacy for different teachers. This chapter

concludes with a section that reports on the contributors to the development of summative assessment literacy identified by the beginner teachers.

While professional knowledge relating to summative assessment is an important element of assessment literacy (Stiggins, 1995), knowledge itself is not the sole characteristic of summative assessment literacy. Summative assessment decision making is informed by knowledge within particular contexts and is influenced by a number of factors, both internal and external to the teacher. Chapter 7 presents findings that relate to the influences on beginner teacher decision making about summative assessment, termed amplifiers and filters. Four amplifiers and filters were identified and are described. Firstly influences external to beginner teachers are described, that is, the sociocultural context which includes education policy, school summative assessment culture, qualifications system as well as students' cultural and ethnic backgrounds. Following this, evidence of the role of influences internal to the beginner teachers are presented. The key conceptions and beliefs teachers articulated that influenced their decision making are outlined, followed by a section on ethics and fairness. Ethics and fairness has been chosen for special attention given the teachers' strong views on fairness and social justice. The final section relates to teacher emotional responses as an amplifier/filter, and describes the effects of beginner teachers' emotions when assessing others and when being assessed themselves, as these acted to guide their summative assessment decision making.

Firstly, this chapter presents dimensions of teacher professional knowledge identified from the analysis of data from eight beginner teachers over the period of the study. Its purpose is to show the overall breadth of beginner teacher knowledge about summative assessment, rather than to focus on individuals. These dimensions are reported under five main headings: knowledge of purposes of summative assessments (Section 5.1), knowledge of what to assess (Section 5.2), knowledge of summative assessment strategies and task design (Section 5.3), knowledge of summative assessment interpretation (Section 5.4), and knowledge of quality summative assessment (Section 5.5).

5.1 Knowledge of purposes of summative assessment

All beginner teachers in this study were able to display knowledge of a wide range of purposes of summative assessment. They acknowledged that summative assessment could serve a range of purposes for both students and teachers. These purposes were informing teacher pedagogical decision making (Section 5.1.1), streaming classes and option selection (Section 5.1.2), reporting to parents and other stakeholders (Section 5.1.3), contributing to NCEA qualifications (Section 5.1.4), affecting student motivation and behaviour (Section 5.1.5), and enabling student learning (Section 5.1.6).

5.1.1 Informing teacher pedagogical decision making

Every beginner teacher viewed summative assessment as an essential element of teaching. Without some sort of summative assessment the teachers asserted they would not really know whether students had learnt what they had been teaching them. They concurred that the fundamental purpose of assessment was to find out what students knew or could do, with the key purpose of informing teacher decision making around pedagogy. All agreed that it was important to do more than just record marks or grades from assessments, and they described a number of useful pedagogical responses to summative assessment with respect to individual classes: changing or adapting future teaching, providing feedback of various types, preparing students with better assessment taking strategies, and helping teachers critique and improve assessment tasks. They also discussed their ability to use aggregated data to inform pedagogical decisions, and as a means to reflect on their effectiveness of their pedagogy, hence informing their future professional development.

Based on their reflections on summative assessment results, the beginner teachers considered they could use the information to identify students' areas of weakness and strengths in order to make judgements about past teaching actions and decisions about future teaching actions. Benjamin and Bella explained:

> Well if you just saw that they got an Achieved [in an NCEA practise task], that they *just* passed, then you know to emphasise something more. [Bella, I1]

> For my own purposes I guess, for seeing whether something worked well or didn't work well. To see if I need to teach something in a new

style. It might be something I could follow up on that and see what actually worked well with the students. [Benjamin, I5]

It was clear from these and other comments that the beginner teachers considered summative assessment to be useful with respect to teaching actions. They talked about deciding to change their pedagogical approach to re-teach or re-emphasise important content or skills when they saw areas of weakness through the use of summative assessment:

> So if they haven't met the specific criteria then you have to obviously, since you've assessed them, you have to figure out how to fix it. So what they haven't figured out. So if you've assessed yourself and what you've taught you have to be able to identify things to change as well. [Elisa, I1]

> I would look over the mock exams and the end of year exams, if you get to see them, and see where the struggles were. And then make that a priority next year to teach it better. Organise more personal reflection like, 'obviously that did not get taught as well as I could have, what can I change? What did work? What can I bring and to make it more interesting? Or more relevant for the students to learn?' [Mary, I4]

These quotes demonstrate that personal reflection, or assessment of themselves, was identified as an important aspect of summative assessment, and that this reflection could lead to better informed future actions.

The beginner teachers explained that summative assessment by way of school-based formal examinations were often held for Years 11–13 students, mid-year. These examinations provided practice for the external NCEA examinations held at the end of the school year, and were also used for reporting achievement to parents. Based on the results of these examinations, beginner teachers explained that they could make decisions about their ongoing teaching in order to better prepare their students for the external end-of-year examinations. For example, after reviewing one summative assessment task, Benjamin planned additional lessons for the students as he decided that more teaching was required by his Year 11 class.

Feedback and feed forward to students based on summative assessments was another aspect of pedagogy linked to summative assessment by all of the beginner teachers. They explained a variety of strategies that they used to provide feedback

and feed forward. In the following quote, Wiremu explained the importance of the action of feed forward six months into his first year of teaching:

> Giving good feed forward is essential. We would not have made the progress we made last year without really good quality feedback and constructive criticism. And recognising that that is quite hard to do, so you plan to do it while you're grading and marking. And then when you give them out and they read 'you almost got that buddy, all you need to do is this, this, and this.' I find that they really respond to that. [Wiremu, I5]

Seven of the eight beginner teachers mentioned specific feedback strategies including going over test questions, giving praise, showing exemplars to students and clarifying ideas through discussion. Benjamin explained his increased understanding of the use of feedback and feed forward at his end of the ITE year interview:

> We sort of broke it down bit by bit because they were a smaller class and I actually gave them a lot more feedback about their results ... Lots of emphasis on the positive rather than tick this, tick this, tick this. I suppose we've learnt the whole feedback/ feed forward thing ... I'd never heard of feed forward until I came into this year, which I thought was very cool. [Benjamin, I4]

Elisa felt providing timely feedback was really important, as it reduced the number of misconceptions students might be holding or developing:

> If it's wrong then obviously you're going to have to address that. And so it kind of makes you clarify things as you go, rather than figuring out it's wrong way down the track. [Elisa, I1]

Six of the eight beginner teachers talked about the purpose of feedback to alert students to assessment-sitting strategy. This sort of feedback was designed to help students approach their next summative assessment opportunities more strategically. Mary explained:

> Practice exams, kind of like mock exams to get them ready, to show them what they actually don't know, what they need to know, where they need to work on, so I guess it's a way of showing students what they need to learn. By grading them on something. [Mary, I1]

For example, subsequent to school-based examinations, Susie gave students feedback about examination technique that they could use in their end-of-year

NZQA external examinations. Benjamin and Mary also explained how students could use summative assessment tasks formatively to direct their own learning and preparation for their final attempts at standards, another important purpose of summative assessment in the NCEA system.

Another pedagogical action that followed summative assessment in some cases was to critique the design of the assessment tasks used, especially when teachers realised the tasks did not give their students sufficient opportunity to fully demonstrate their learning. Two teachers spoke of how the experience of assessing their classes caused them to redesign the assessment tasks before using them again. They gave examples of tasks they had used that had not given their students sufficient opportunity to provide evidence of their learning, or had not matched learning outcomes adequately.

By the middle of their first year of employment all of the beginner teachers felt they could attempt to analyse and utilise summative assessment data and understood the importance of doing this to inform ongoing pedagogy. Wiremu compared the analysis of summative assessment data to scientific method:

> Well, yes, if you gather data without analysing it, you've only done half the work haven't you? … You don't just get the data, and just be happy with it. I do think like a scientist, so I think scientists have a little bit more of an advantage in that area … You need to analyse the results, and write up the analysis of your results. Even that is not enough. What do you do once you have results? Conclusions, discussions, what further questions can be asked? And so just to foster that scientific mind, and just to not lose it by being the teacher, by worrying about being the teacher. Because there is good data out there and who better than science teachers to analyse the data about science in schools? Why wouldn't I be the authority on that? [Wiremu, I4]

Using skills from their science backgrounds and being able to think like a scientist, most teachers felt somewhat comfortable with the idea of analysing data at a basic level. They talked about having mathematics skills to do basic analysis so they could reflect on their students' progress and potentially identify anomalies, or areas for further investigation or action.

However, the beginner teachers admitted to having very limited knowledge of how to analyse or use aggregated summative assessment data for pedagogical purposes. They were keen to use students' results as a basis for reflection on their

teaching over time when thinking ahead to the next year's programme and for planning. Five teachers explained that they had not been given any explanations about how data could be aggregated and analysed or used by teachers, and that they only really focused on individual student achievement.

> I don't think so. I don't think teachers do a good job at collating students' test results to see how well they have done … I don't think a lot of teachers do anything with it. They probably aren't taught that. [Mary, I4]

Mary felt very underprepared in this aspect of her teacher education, as she did not know what she could do with students' aggregated results. She saw no evidence of data analysis at her practicum schools.

Three teachers did get a little insight into how aggregated class data could be analysed. In this example, Kate talks about one of her ATs being helpful in this regard:

> But with my Year 13 form teacher, she was in digital technology and was quite digital savvy and we were going through KMAR [student database] and she was showing me a few things and I just clicked on one thing randomly and it came up with results and things like that. And being able to use results to differentiate the way you should be teaching, because the result is a physical result, and then it is how can we expand? … What should we change? [Kate, I4]

Kate explained how she could look at class averages, and at the results of subsets of her class to see performances in individual science topics, and use other data, such as social groups, gender, literacy levels, numeracy levels and previous years' grades to monitor progress and help inform her practice. Some of the artefacts Bella produced were Excel spreadsheets that showed aggregated data from her Year 10 classes. Ryan was the only teacher who had received specific professional learning about data literacy during the study, as he was sent on a one-day data literacy course which he found informative.

The beginner teachers also spoke of using results more generally as an indicator of their own progress, using them in critique of their pedagogical decisions. In the following quotes, Kate and Susie talked about what they could gain from reflecting on summative assessment results:

145

And I think it's a way of formulating assessment to know what the students know or have learnt, and also as a way to direct back at yourself as to what you can change. [Kate, I1]

So I was just doing internal assessments, so it is just getting grades, but it is also how you get feedback. It is kind of your way of getting feedback on your teaching as well. [Susie, I4]

This feedback helped them to critique the effects of their pedagogical decisions, and to make decisions about their own professional learning needs.

In summary, there are different foci for decision making based on summative assessment. The beginner teachers talked about its use in planning or adapting future teaching, and informing their pedagogical decision making for individuals or classes, although there was less certainty and confidence across the group in the use of aggregated summative assessment data. All teachers also valued the ability to critique their pedagogical practices through their reflection on summative assessment results.

5.1.2   Streaming classes and option selection

The beginner teachers in this study identified that another purpose of summative assessment was to inform school-level decision making related to school structure and systems such as for streaming classes, for Head of Department reviews of classes, programmes and staff effectiveness, and for school comparisons and national rankings. This was explained by two teachers as follows:

I'd say Head of Departments would … use it to see how different classes are going, and maybe different teachers and teaching styles. And then in a broader national level to see how the nation is going. I'd say also future teachers are better looking back on past assessments on classes that they're going to teach the students, and if students transferred from other schools, to see where they're at. [Mary, I1]

It serves different purposes for the students, the teachers taking them, and the people doing sort of organisational things. Like the classes are streamed into "one band", and "two band", and "three band" … because I guess with assessments that's the whole thing, its decisions need to be made and you can't just make them willy-nilly. You've got to have something to base them on … So that's why you need that assessment. [Wiremu, I2]

In these quotes Wiremu draws attention to the need for robust summative assessment from which to make streaming decisions, rather decisions being made

in a haphazard or "willy-nilly" way. Mary's comments highlight the broader uses of summative assessment by senior staff members and others, both within a school to ascertain class progress and teacher effectiveness, and to allow external comparisons to be made.

Benjamin commented on the way physics teachers used grades to decide student entry to the Year 13 physics class:

> Well, streaming for next year was always emphasised. I think that's used as the most … not blackmail, but something to get the kids to knuckle down… like for the top level physics class, there is only one top-level class and they do mention that a hell of a lot. So the summative grades are used for that. [Benjamin, I4]

In classes with restricted entry, such as physics, students were forced to compete for places and decisions were based on grades. Benjamin observed that summative assessment acted to provide extrinsic motivation for students as well as to enable teachers to decide on class composition. Streaming classes and option selection was identified as a purpose of summative assessment, but usually the responsibility of senior teachers.

5.1.3   Reporting to parents and other stakeholders

Formal reporting of student achievement was another purpose of summative assessment that the beginner teachers identified. The grades resulting from school examinations, NCEA standards and other assessments offered by schools are documented in school reports. All of the teachers demonstrated knowledge of the requirements and processes of reporting learning progress. However, this was not viewed as the primary purpose of summative assessment by any of the teachers.

Three of the teachers questioned the narrowness of what was reported on by schools because although it reflected what was assessed by way of subject content, it ignored other aspects of students' knowledge, skills and development. For example, Benjamin was interested in helping students develop thinking skills, but the only way to report on these was via a comment on student reports, as there were no tests or grades allocated for this skill. As illustrated in the quote below, Benjamin felt that parents pay more attention to the test marks and grades, as they see these as a more robust measure of their children's achievement than a teacher's comment on skill development:

> I know which students are good at [thinking skills] and which are not, and I can praise them, but it's still not really being recognised like within their reports or a formal setting. I guess next year I could do the same thing, and I could put in their reports or tell their parents that they did well, but then because it is not really like a test … parents will probably look at the test marks and think that that's verified, and then you might put in his report that he's really creative with good thinking, but I'm not sure that if they've got poor test marks that if that'll count for much. [Benjamin, I4]

In contrast, Mary had learnt from her AT on practicum that regular recording of non-tested attributes of students was useful when it came to report-writing time. She valued her AT's use of a spreadsheet on which regular notes were made on issues to do with learning as well as homework completion, having required equipment, and so on. This full record helped her AT write reports. She followed a similar pattern when she started teaching and said her regular notes reduced her stress levels when it came to report-writing time as all the information she needed was summarised on the spreadsheets.

Elisa acknowledged that report comments could be used to put pressure on students to take their school work seriously, as they generally do want their parents to read positive reports:

> … and so I think when they've got the pressure on, like "This is on your report, this is what you've got to know for NCEA", and they're like "Uh-oh, this is serious." [Elisa, I2]

Four of the teachers in the study were able to contribute to student reports and to parent-teacher interviews while on practicum; all were involved in report writing once they were working in schools. They all said that practicum helped their understanding of this purpose of summative assessment. All teachers were involved in writing reports in their first teaching positions.

Overall, the beginner teachers understood the requirement to report on student progress, but they did not view it as a primary purpose of summative assessment.

5.1.4   Contributing to NCEA qualifications

A key purpose of summative assessment, identified by all beginner teachers in this study, was to enable senior students to gain credits for NCEA. All teachers recognised this purpose and took seriously their role as assessors within the high

stakes NCEA system. Because of the importance of NCEA they felt considerable responsibility to their students, and all were concerned about developing their knowledge for this purpose of summative assessment. Kate and Mary's quotes are representative when they both talked of the fear they felt if senior students did not do well in their classes:

> Because I'm scared that I won't be hard enough on the kids with those expectations, and so we will drift through seven weeks and then realise. [Kate, I4]

> I guess that is the biggest fear, it's just, oh man, imagine if I did this whole test and they all failed. I suck as a teacher … When there are 15 kids in the class and four of them don't show up … So that doesn't look really good for our record does it? [Mary, I5]

All of the teachers were aware of their need to learn as much as they could about NCEA assessment, and they appreciated what they could learn through university papers, as represented in this quote by Bella:

> Yes it was good when [a lecturer] would give us lots of Standards, and would talk through them, the teacher guidelines as well as the student guidelines. Knowing exactly what we had to do and putting it into context as well was really helpful. [Bella, I4]

In summary, the beginner teachers had varying views on standards-based assessment but all knew about the critical role of the teacher in the NCEA system, contributing to students' qualifications. The individual elements of NCEA assessment are covered in later sections in this chapter.

5.1.5 Affecting student motivation and behaviour

Student motivation and behaviour management both featured in teacher comments. Five out of the eight teachers in this study contended that enhancing student motivation is one purpose of summative assessment. This point has already been raised in relation to class entry and reporting, be here it relates to student learning behaviours. The comment from Mary was representative of this view:

> Maybe assessment is a way to motivate them. [Mary, I3]

In particular, summative assessment tasks leading to NCEA credits were seen to provide motivation for students to push themselves to succeed or achieve a goal. Elisa illustrated this point:

> So, like, if they're going to work towards an assessment then they are going to do the work. If they don't have a goal … then they are more likely to not do it. [Elisa, I1]

However, all of the teachers observed that such behaviour almost always seemed to lead to students developing a "credit-grabbing mentality" [Ryan, I2], which was seen to cause students to lose motivation once they had gathered the minimum required credits for the NCEA certificate they were working towards. They were critical of schools which over-emphasised credit-gathering ahead of learning, as exemplified by Mary:

> Why are we learning this? I think when it gets to senior school some of them don't need credits and so that motivation isn't clear. [Mary, I4]

All the beginner teachers felt that, in general, the structure of NCEA caused students to be focused on the accumulation of credits rather than deep learning of science:

> Yes, I felt like teaching was definitely for credits, and that's what the students were all concerned about. They were just sitting in class counting up how many credits they got … and they didn't care what they were doing. So it's just discouraging … Because they could all get Excellence on the one I was doing because it wasn't hard. But they are all like, oh, what I do care about it, all I want to do is pass, how do I pass? [Mary, I2]

Mary identified students who were credit-counting, focused on the accumulation of credits rather than reaching the excellence level grades of which they were capable, and she found this lack of student motivation disappointing.

When discussing the effect of summative assessment on motivation for junior students not enrolled for NCEA, the teachers acknowledged that upcoming summative assessment tasks were less likely to motivate students to study harder because the stakes were much lower for them:

It's almost harder to motivate them [Year 9 and10 classes] I think because they know it's not always going towards anything. It's just like a letter on my report. [Bella, I1]

It emerged that the beginner teachers in this study were more focused on using informal summative assessment strategies for junior students to try to keep their junior students engaged in science, rather than using robust summative assessment tools to ascertain their levels of understanding:

All you want them to do is make them enjoy the subject. So it didn't really matter if I gave them a really easy test so they would get good grades and feel better after that. [Elisa, I4]

Elisa talked about making sure the content and design of summative assessment tasks for junior classes was pitched at a level that would ensure student success, as she felt attaining success was more important for junior students than being able to rigorously report on learning. All of the teachers were concerned about encouraging junior students to enjoy science and think positively about it so that they would continue studying science subjects in senior classes in future years. They believed that success in science assessment would encourage this.

I think in the juniors it is important to make it more fun and exciting, not just "this is the physical word" "this is a material world" … But to merge them with activities and stuff. Because I think that's how they get interested in science. [Bella, I4]

Bella, Elisa and Susie provided artefacts that confirmed that they were trying to make their summative assessment tasks fun and motivating, through the use of cartoons in tasks and by giving students choice.

A number of beginner teachers also talked about using summative assessments as a way of controlling or managing the classroom and homework behaviour of students. Benjamin explained the use of assessment as a means to make sure they completed homework, rather than on the assessment of learning, although in his mind this was a pointless use of assessment. With respect to classroom management, Bella commented:

Yes, I guess if you have a really loud class and you can't manage them, you can just give them a test or something to keep them quiet. [Bella, I3]

This purpose of assessment extended to teachers' marking and grading, as teachers indicated that being 'tough' rather than 'soft' when marking exams also helped with classroom management. When students gained lower marks one teacher maintained this would spur them on to work harder in class.

So although the teachers in this study did not approve of using summative assessment to manage student behaviour, a number confirmed that this was one of the purposes they had knowledge of and could be one they might employ.

5.1.6 Enabling student learning

Three beginner teachers in this study described the benefits of summative assessment for learning. They explained that summative assessment tasks themselves could promote learning if they were designed well, as they provided learners with opportunities to draw their science knowledge together and synthesise what they had learnt. This is a feature of some internal NCEA assessment tasks which require students to draw upon a range of science knowledge as Susie explained. Susie described a biology standard which required students to integrate biological knowledge to develop an informed response to a socio-scientific issue. For this assessment students were able to draw from a range of topics they had learnt in biology. The two teachers who had assessed against such standards were very positive about the benefits such summative assessments had for learning:

> But in the couple of internals that I did, I did the biotechnology in
> Year 13, and it was a report, and it ended up being really good just
> doing the report writing … it was a way of pulling all of the ideas
> together for the students as well … so they were asking very
> intelligent questions, and there was lots of very big picture stuff that
> they were thinking about in their research … Because you know
> science is taught in units but it relates so much to everything else,
> nothing is on its own. [Susie, I4]

Three teachers provided examples of assessment tasks they had designed to draw on knowledge from a number of science topics. They saw the skills and synthesis of knowledge required of this type of science summative assessment task as being more like the reasoning 'real' scientists use. They also commented that because science is taught as topics, it is often not until students strike a complex

summative assessment task that they are forced to consider the contributions and interactions between a range of science topics at once.

In summary, Section 5.1 has described beginner teachers' perceptions of summative assessment purposes. These purposes served students, teachers and other stakeholders. They were categorised as informing teacher pedagogical decision making, streaming classes and option selection, reporting to parents and other stakeholders, contributing to NCEA qualifications, affecting student motivation and behaviour, and enabling student learning. The following section reports on findings on teachers' knowledge of what to assess when using summative assessment.

5.2   Knowledge of what to assess

A second category of knowledge of summative assessment displayed by beginner teachers in the study was knowledge of what to assess. The teachers talked about knowledge related to science content as concepts and processes (Section 5.2.1), and content related to wider curricular goals (Section 5.2.2). Their knowledge of what to assess also included important skills and dispositions not necessarily directly linked to science content (Section 5.2.3).

5.2.1   Science content as concepts and processes
In New Zealand, teachers must teach aspects of physics, chemistry, biology, earth science and astronomy to Years 9–10 and then either general science or specialist science subjects to senior classes (Years 11–13). All beginner teachers in this study believed they needed a thorough understanding of the concepts and processes involved in science topics they had to teach. They knew they had to assess science concepts and processes. Scientific knowledge was understood by all of the beginner teachers in quite a traditional sense as facts with scientific knowledgeseen as relatively stable.

Although they were all science graduates, when they were interviewed six months into their first teaching roles none of the teachers felt entirely comfortable having to teach and assess the breadth of science content they were expected to teach. They were being expected to teach science topics that they had not learnt about at university or in some cases had not learnt as part of their university studies. Bella commented,

> My associate helped me out for the sequence which was quite
> helpful because I don't think I learnt astronomy at school. [Bella, I1]

The teachers considered they needed to have a sound understanding of science topics themselves if they were going to assess them. Like the others, Bella's first response when asked what she wanted to make sure she had "under her belt" before she started assessing students, was to have a thorough knowledge of the relevant science content:

> I guess my own knowledge of the topic that they have to know. [Bella, I1]

Their perceived lack of breadth of science knowledge worried them, particularly at the beginning of the course, as Ryan explained:

> I worry a bit about my content knowledge … I'm going to have to
> revise a lot before I'm ready to evaluate someone else's performance.
> So on that basis I'd say no, I'm not really ready. [Ryan, I1]

Ryan commented that the workload generated by ITE requirements meant there was little time for him to revise and learn the unfamiliar science content he was having to teach on practicum:

> My content knowledge is not sufficient, and I know that. I have
> known that from the outset. But you don't have the opportunity to lift
> it because you're too busy with all of the other work. [Ryan, I3]

Ryan felt this put him at a disadvantage when teaching and assessing students when on practicum and when he started teaching.

In topic areas where the beginner teachers felt confident about their science knowledge they all felt more able to make summative judgements about their students' work. They linked their having the background knowledge of the science content to their ability to understand student thought processes:

> Well, I've got the background knowledge of the subject, so I'll be able
> to assess. I will have a formal understanding of the thought processes
> as they are doing it, and the assessment. [Benjamin, I1]

On the other hand, there were times when the teachers knew their subject content very well, but considered they lacked knowledge when it came to preparing their teaching and assessment materials at the right curriculum level for their secondary

students. They sometimes struggled to reduce complex university level knowledge to the level required at secondary school. Susie explained the challenges she found in teaching and then assessing complex concepts in a very limited timeframe in the following quotation:

> But it was insane … Like, biotechnology is a completely new topic to them, they haven't done genes since Year 12 … it's like they have some ideas with speciation, and a lot of the ideas all come together. But all the technology is brand-new and I had a week to teach transgenesis, a week to teach selective breeding, and then they had to do their own revision. And there were 13 lessons including the week of research and the week of write up of the report, to teach a topic ... I took four years learning that stuff. [Susie, I4]

Artefacts that Susie produced showed that she had thought carefully about how to pare down the science content so that it was not overwhelming for her students. Her assessment task design allowed for student learning as they completed their biotechnology reports.

All teachers in this study were very appreciative of the topic specific content knowledge they gained as a result of the science curriculum courses in their ITE programme. The science curriculum workshops in particular helped them link science content to appropriate learning outcomes and curriculum levels. This gave them confidence, as illustrated in these quotes:

> For Year 11 science we were doing chemistry and doing fair testing, so a lot of the stuff that we had done with [lecturer's name] … It was really good just knowing all the details and feeling really comfortable and confident about knowing how to do a graph. [Susie, I2]

> Some of the things with [lecturer's name], like the particle nature of matter. I was quite stoked when I was doing that topic [on practicum] … So a lot of her stuff was really handy. [Bella, I2]

Having planned some science units in the university ITE course gave these teachers more confidence when they came to teach and assess the same topics. For example, Kate and Bella explained that while on practicum they were able to use a series of lessons that had been developed in an ITE workshop under the guidance of a university lecturer. They both produced copies of summative assessment tasks they developed for the unit, with which they were very happy.

All the teachers wanted to continue to improve in their knowledge of what to teach and assess, and to become more aware of how to deliver learning at the appropriate level for their secondary school learners and then assess it appropriately. Kate's comment sums up this point:

> I think certainly university has done that with [university lecturer's name], and the things we could achieve in a lesson. That's been able to help me know how to better pitch it. [Kate, I2]

A number of the teachers felt that even more of a focus on the teaching of specific science topics in the ITE programme would have been beneficial.

In summary, the beginner teachers described their science content knowledge and their understanding and desire to keep developing knowledge of science, including how to "pitch it", or decide on the content levels best suited to classes. Work done within their ITE programme helped with their knowledge and confidence in assessing specific science content.

5.2.2   Content related to wider curricular goals

As well as the summative assessment of science concepts and processes, the beginner teachers in this study understood that they needed to summatively assess content related to wider curricular goals. They linked the content of a topic to the practical skills used in science (such as graphing, report writing), the practical application of science knowledge in society, and notions of "thinking like a scientist" as well as to their own knowledge of the requirements of standards they were going to use. These were seen as being integral parts of the subject of science. A number of the teachers thought it would be important to teach and assess these aspects of students' learning alongside scientific facts:

> We need to teach about how to write a report, and teaching about APA referencing and things like that ... Year 11 students didn't know how to graph. And I didn't realise that I need to teach them in science how to graph, because in coming in from economics and maths where they all graph differently to what we want in science … Okay they actually need to have these things for assessments. [Mary, I3]

When talking about their knowledge of what they should assess for specific year levels, the teachers emphasised the importance of understanding the requirements of the ASs and USs for NCEA. These standards told them what to assess in broad

terms, and the supplementary resources helped them understand the requirements more specifically.

> I guess with the Microbiology unit, the internal [achievement standard], seeing the standard and knowing what they actually have to get to pass an achieved, merit and excellence, although you don't really want to teach to the standard … So that kind of helped, the 'going backwards'. So it was really good working entirely through it. [Bella, I4]

> I designed the Year 13 Biotech one, but you are still kind of set within guidelines, you had to do a report, you had to research. [Susie, I4]

Although teachers admitted concerns about their lack of content knowledge for some topics, none of the teachers referred to this as an issue when they marked NCEA assessment tasks. This was because they prioritised learning the requisite content required for the achievement standards by the time the assessments were completed because of the high stakes nature of these assessment tasks and the accountability measures in place.

5.2.3   Other important skills and dispositions

Beginner teachers in this study acknowledged that in their classrooms students learnt much more than science content. They emphasised the need to assess other aspects of learning such as thinking skills and curiosity. For example, Benjamin saw the NZC as a document that espoused ideas such as creativity, forward thinking, problem solving, critical thinking and so on, but he felt that these broader learning goals were not assessed by many teachers:

> I think their thinking should be assessed. It would be hard to assess, but give them like a foreign situation and see what their thought processes would be. Being creative and thinking, because they always say 'encourage creativity in this or that', but creativity is not really assessed, I don't think. Well for maths and science, and those are probably two subjects where you definitely need creativity. Like problem-solving. [Benjamin, I2]

As well as this, Benjamin voiced concern over the fact that skills required for successful integration of students into the workforce were overlooked in school assessment. Benjamin felt that these broader skills and knowledge should also be part of what is assessed. So he saw what should be assessed as being much broader than science content:

> Because I think at the end of the day we are training them for the workforce as well, and I think when you go to job interviews, because I've done engineering as well, they always wanted people who could get on and work by themselves autonomously, but then also they could react to foreign situations where they could be creative and unique. I don't really think that's being assessed, like we encourage it [but] then it is not being assessed. I think there should be more emphasis on that. [Benjamin, I1]

Another important disposition that teachers felt should be assessed was curiosity and questioning. Wiremu explained that one of his priorities in science teaching was to encourage students to be curious and ask questions. He saw this disposition to be as important as the actual science knowledge that students were gaining and believed it should be assessed.

> Because I think half, at least half of science, is the enthusiasm and curiosity of wanting to ask a question. So I think even if the kid has started quite low [academically] in their science capabilities, and you know they are not very enthusiastic about their science capabilities but by the end of a couple of terms their grades may not have picked up, but their classroom participation and their enthusiasm for science has gone up, then in my books I'd see that as a personal assessment as their science skill has increased. Not in books or anything, because half of science is to want to ask questions ... for me not just in the grades but in the enthusiasm for science. I think that is another way that you could just tacitly judge and assess them on. [Wiremu, I1]

In this quote Wiremu endorses the assessment of students' disposition towards science, and emphasises the curiosity and question asking central to scientific investigation.

There were some surprising omissions in teachers' responses about their knowledge of what to assess. None of the teachers in this study mentioned the need to identify and assess the central or key science concepts that students need to master because it is on these that other science knowledge is built (e.g., the significant need for students to understand atomic structure). Instead, all science knowledge appeared to be regarded as equally relevant for assessment. Additionally, some very important elements of the NZC were scarcely mentioned by teachers. Key competencies were only mentioned by one beginner teacher once and this was not with respect to assessment. Four of the teachers mentioned that they had learnt about the nature of science at university but only one talked about integrating this into his teaching. None of the teachers talked about the need

to assess students' understanding of the nature of science, either formally or informally. Although students' involvement in science inquiry was acknowledged, it seemed that teachers in this study did not know about assessing students' understanding of the inquiry process. One teacher did talk about the need for students to be able to ask good questions in science but did not follow this up by talking about how the questions students raise could feed into an inquiry process. It seemed that for almost all of the units in which inquiry was a factor, teachers provided students with a supported and scaffolded inquiry model, meaning the students were not expected to develop and exercise independent inquiry skills and nor were they assessed on these.

In summary, this section has reported on the teachers' knowledge of what to assess, both science-specific context and students' broader skills and knowledge. Teachers were able to talk and demonstrate their understandings, but also articulated their lack of knowledge in some areas. Knowledge of what to assess is critical when designing or making choices about assessment tasks. The following section reports on teachers' knowledge of assessment strategies and designs.

5.3 Knowledge of summative assessment strategies and task design

The third category of knowledge of summative assessment was teachers' knowledge of summative assessment strategies and task design. This section presents findings about teachers' knowledge of a range of strategies (Section 5.3.1), and assessment task design (Section 5.3.2). In this section a summative assessment strategy or task refers to the instructions, task and marking criteria associated with an assessment event.

5.3.1 Range of strategies

The beginner teachers demonstrated knowledge of a wide range of summative assessment strategies and designs. They all talked in considerable detail about the use of tests, examinations, collaborative group tasks, field work, written reports, presentations, observations, posters, rubrics, and so on to assess student learning. Over the course of the interviews they each explained their own use of some of these strategies and talked about their observations of strategies being used in the classrooms of other teachers. They spoke about their experiences using various

strategies in a very reflective manner, often discussing their perceptions of strengths and weaknesses of particular strategies, and were keen to learn more. Although all of the teachers were able to talk about the use of summative assessment strategies, it was clear that some had much more exposure to using a range of strategies than others. Nonetheless, they all articulated how their confidence in using different assessment strategies had developed considerably over time. They were all keen to try new strategies, including digital tools. This breadth of insight is reflected in the following teacher quotes, both from the second interview which occurred after their first practicum experience:

> Like, when we get into class, and I've got a Year 12 Bio class, and they're going to be doing Adaptations, and I am quite familiar with that from doing physiology all through uni, but learning what are the standards? What are the standards for year levels that I should be tailoring my lessons to being able to answer, and how deep do you need to go? And so I guess just knowing which kinds of assessment techniques are really relevant to science and biology. I've got a few, but just keep asking, picking the brains of our experts we've got here. [Wiremu, I2]

> It's interesting, the classes that I did take where there were practicals and I did let the students take photos of the process ... when you don't have time it is easy to slip back into being the traditional teacher, like you need to do a test so let's just do this test … multiple-choice, write answers, that's the test… Rather though I think there could be more creative ways of testing people … I want to use some different creative ways. In our class yesterday they were talking about using texting polls, or ways to get students' feedback, which would be interesting to use. [Mary, I2].

As these quotes illustrate the teachers felt challenged when trying to choose the best assessment strategy to assess a particular topic, but they had ideas to draw on.

All teachers in this study were very aware of the benefits of using a broader range of assessment strategies for students rather than relying on traditional tests and examinations. Two teachers expressed the following views, promoting the use of a range of strategies both to allow students to better represent what they have learnt, and to help them learn:

> The assessment type is very specific, depending ... And so in that sense you need a wide range of understanding of assessment so that you can then make it the most effective way of learning. [Susie, I1]

> I think some students don't test well, and so you don't get a true representation of what they have learned. The good students are always going to score good and the bad students [are] always going to score bad, but the students in the middle … It's a bit of a blurry line about whether they actually know it, or whether they are being lazy. It's about using a range of assessments tools is a key. [Mary, I4]

The teachers also understood the benefit inherent in matching assessment strategies with students' strengths and interests. When they could do this they felt that the results gave a fairer representation of the students' learning. For example, Susie provided examples of assessment tasks she had designed for Year 9 students, where students could choose the style of assessment task they wanted to use, whether a poster, or a web based quiz, or an oral quiz. She found that most of the students chose the poster option, which she said matched their interest in visual art. Susie felt that by giving the students the option of assessment type, they were more likely to produce their best science work.

All teachers in this study discussed formal strategies often involving planned written tasks, and informal strategies (more likely to involve observation or chatting during the normal run of the lesson). They conceded that in secondary schools, formal summative assessment plays a big role, but three teachers spoke of the use of teacher observation and other informal means as being better ways to assess students summatively. They felt they were able to make better judgements after engaging in discussions with individual students, rather than depending on the results of formal written assessment tasks:

> In that sense the kids can tell you a lot about how things are progressing, both by the questions they ask, and the answers they give to your questions. I actually found I preferred doing that a lot more than actual tests. [Wiremu, I2]

The teachers understood the need for formal assessment, but valued informal strategies and felt that these often allowed for better summative assessment.

5.3.2 Assessment task design

The beginner teachers expressed concern about their level of knowledge of assessment task design, feeling underprepared to design assessment tasks at the end of the ITE course. They generally appeared tentative over the flexibility of assessment design (including ideas of relating to co-construction), and what was

permitted with respect to NCEA. Knowledge with respect to literacy demands in summative assessment was discussed, as was their understandings of effective and efficient tasks that promoted learning. All the teachers rated their knowledge of the range of possibilities of assessment task design as limited.

Effective task design was seen as difficult and complex by all the beginner teachers, as explained by Wiremu:

> My whole issue with tests was that it doesn't necessarily gauge how much learning has gone on. So from that point of view from now on I realised that creating assessments must be so difficult because I guess those who create them are aware of that, and they aware of that you know you can't make the wording too easy yet at the same time if they get caught up on how to read the question then that is just an issue in itself. [Wiremu, I2]

Teachers had a variety of views on how much flexibility they had in designing summative assessment:

> I thought there was more of a set out curriculum that everyone does, and I think depending on the school, there would be more of a structure. But it seems as though you can write your own assessments up until NCEA, where they (the students) actually have to know some things. [Mary, I1]

> It's just kind of having those options when designing it, for the students in that class too … So it's just getting to know your students, and designing your assessments around the students. You still need to follow the guidelines of the rest of the class, but adapting it slightly. [Susie, I4]

As these quotes illustrate, some teachers considered that they had some flexibility in assessment design, particularly when it came to issues of equity, whereas most felt more tentative.

All of the beginner teachers expressed uncertainty about what and whether there was a required way to assess the NCEA standards. They all stated that they did not understand fully what was permitted and expected within the NCEA system:

> Maybe [what I need to learn is] just further ways to assess because we still haven't learnt a lot about assessment, so I'm not sure. I'm confident that I know I am assessing them, and assessing their learning, but I'm not sure what the 'given way' of how we are supposed to be assessing is. I'm not sure if you're supposed [to] use a

given way ... yes that's something I'd like to work on. Just different ways to assess I guess. [Benjamin, I2]

I didn't really understand who gets what grade, and who can re-sit what. So it seems like there is a lot of technical stuff about NCEA and I don't really know what's going on. [Mary, I2]

For example, Bella produced artefacts which she felt represented her ability to assess life processes at the cellular level with a Year 12 NCEA class. Her summative assessment task was a compilation of test questions used by a number of teachers in previous years. She felt more comfortable using what could be seen as a conservative test format for this assessment even though her planned unit contained a number of innovative teaching strategies and resources. Similarly, Ryan's assessment of genetic variation for a Year 11 class was in the form of a test using NCEA terminology, despite evidence in his unit planning of innovative thinking that could have led to innovative assessment.

The teachers appeared very aware that the literacy demands of summative assessment tasks could be a barrier for some students and the need to design tasks in which the literacy demands were not too high for the students being assessed. The three teachers who had worked with students with low literacy levels talked about their need to adjust standard classroom assessments in order to make them accessible to students. They understood that if this adjustment was not made then the assessment results would not reflect their students' learning in science. Wiremu also raised issues that face teachers of first-language Māori speakers when they are considering the language of the assessment in task design. As a bilingual teacher he used both Māori and English when working with those who had transferred from Māori–medium schools into English–medium schools. He had personal preference for oral assessment and he felt that using oral assessment would benefit some Māori students given that for Māori, oracy is a valued aspect of their cultural heritage. It was obvious that Wiremu reflected on his own experience in Māori–medium schooling where oral assessment was more highly valued than it is in English–medium settings, and he linked this to his understanding of assessing students similar to himself:

I still feel that oral assessment should be implemented a lot more … Often there are Māori students who have been in total [Māori] immersion schools, and there is really no strong emphasis on written

163

assessment in total immersion schools … In wharekura [Māori–medium schools] you are also graded on "how well do you speak" and I think that is a huge, huge difference [to English medium schools].… So to say "you have to be assessed in this form" if you prefer the verbal, I think it does hold certain students back. [Wiremu, I4]

Wiremu was particularly interested in designing assessment tasks useful for NCEA science standards that did not require students to read and write. He explained:

… [my] general focuses [for improvement]. The first one is finding as much information and talking to as many people as I can about non-reading and non-writing assessments. What ideas are there out there about assessments that are not reading or writing, what can we do, what needs to be put in place? [Wiremu, I4]

Wiremu and Benjamin talked about their knowledge of assessment design with respect to co-constructing aspects of summative assessment tasks with their classes. They both believed that this provided ways to obtain more accurate evidence of their students' learning, given that the students themselves were more likely to engage with the tasks if they had input into the design of them. For example, Benjamin commented,

Maybe a better assessment might be to let students choose whether they want to give a written essay or give an oral presentation. I think that would be more fair. [Benjamin, I3]

Three of the teachers asserted that strategies that were exciting or fun for the students were more effective than traditional tests. These teachers believed that "fun tasks" equated with a higher level of engagement from students, and hence it would be more likely to get an accurate gauge of their learning. The following comment is indicative of the breadth of ideas raised by the three teachers:

Assessment doesn't have to be just tests and the internals and the externals, but you can have a lot of informal ones, mix and match activities, flowcharts, concept maps and stuff like that. It can be fun. It doesn't have to be boring, and scary and stressful. You can make them exciting, not really the externals, but you can make the internals more exciting as you link them to context so it can be more exciting. [Bella, I4]

However, the beginner teachers observed on practicum that formal assessment types, such as tests and reports, were used by the majority of secondary science

teachers, without any consideration of students' enjoyment. Wiremu was the one teacher who saw this kind of formal testing and examinations as being part of the 'rite of passage' to becoming a scholar. He felt that to be treated as a scholar, students needed to get used to being assessed as a scholar. He saw the academic system as one of rigour with academic progression being marked by achievement through formal assessment processes. So although he preferred oral assessments, he acknowledged the need for formal processes in academic systems.

Project-based assessment tasks were seen as a positive and desirable design type. Using research projects as a vehicle for formal summative assessment was understood as being advantageous for students, and beginner teachers felt it more closely simulated real-life situations. As discussed in Section 5.3.1, these sorts of tasks assessed the students' abilities to apply appropriate science knowledge, rather than simply assessing whether they knew facts and had the added advantage of promoting deeper learning for students, and promoting and drawing together the integrated use of a range of skills and knowledge from a number of science topics. The strategy of holistic project-based assessments was identified as a preferable method to assess a range of skills as well as giving students the opportunity to display their creativity and higher level thinking:

> I think that summative assessment is more effective when done in conjunction with something else, like an interactive activity or something that embeds it more. [Ryan, I3]

> I think the internal [NCEA] assessment where they actually have to do a bit of research … [how] nuclear power plants affect the environment … It's how it relates back to the social, economic, society … I think students have a hell of a lot to learn from that … you want them to be exposed to that … Making links to the big picture and they are also learning to learn for themselves. [Benjamin, I4]

Beginner teachers in this study also talked about the need for efficiencies and design of tasks so that a number of NCEA achievement standards could be assessed simultaneously through larger teaching units or projects. At an aspirational level at least they knew that assessing in this way would be beneficial for learning and would reduce assessment pressure. They felt that broader boundary-crossing assessments more closely resembled real life where ideas are not necessarily blocked into subject areas or topics. Benjamin talked about the

integration of mathematics and physics, for example. Although the teachers generally did not have opportunities to plan integrated units during the course of this study, they made it clear that they would prefer to work in this way. For example, Susie described with enthusiasm a project that included researching the use of edible vaccines in particular third world contexts. Susie was able to experience this integrated approach to teaching and assessment in the school in which she was employed, but others in the study were yet to get the chance to work in this way.

In summary, this section has illustrated beginner teachers' knowledge of summative assessment strategies and designs. The teachers appeared to have a broad knowledge of assessment strategies, but in many instances did not understand the detailed requirements of assessment types, nor did they consider issues to do with assessment task quality to any degree. They were still not confident to design assessment tasks by the end of the study, but did think about students when considering which assessment strategies to use with particular groups of students. The following section reports on teachers' knowledge of interpreting students' evidence of learning in order to make summative assessment judgements.

5.4   Knowledge of summative assessment interpretation

The beginner teachers in this study understood that their teaching role involved the consideration of the evidence produced by students, in the context of identified learning objectives, in order to make judgements using associated success criteria. This section reports on teachers' general knowledge about interpreting and making judgements about student work.

All beginner teachers in this study voiced concern over their limited knowledge of how to interpret students' evidence of learning, and mark or assign grades to the work, and were generally nervous about marking. However, they all did explain that their knowledge and confidence improved over time. By the middle of their first year of teaching the teachers rated their knowledge and confidence much more highly than earlier in the study:

> I found it quite stressful at the start that students' grades were on me.
> But all the marking that I did, [teacher's name] moderated them and

we came out with the same grades. So that was good, it gave me a bit of confidence. [Mary, I5]

Good, I feel capable. Much more solid than I did last year. Actually doing it and having to do it, having to mark it and moderate. It's good. [Bella, I5]

By the end of the study all teachers demonstrated knowledge about the need for clear assessment criteria and a marking scheme in order to be able to mark fairly and accurately. They linked this to their being able to justify their grading, but they still talked about the complexity of the process, especially when developing and applying specific criteria which were based on the more general criteria from achievement or unit standards:

I mean you can't just say I gave him an Achieved. There has got to be a reason behind it … It's like everything has to be based on a criteria and justified somehow. That's quite good to be made explicit in class when you realise these things, and that you're not wrong … because even the criteria itself can be interpreted differently by different teachers. And in that way, by developing your rubric, it is a good pedagogical technique because when you come to justify why you answered it, by using your rubric it might be a lot easier to justify your reasons, than by going back to the general criteria. [Wiremu, I3]

They valued and learnt a lot from the practise tasks given to them in university courses which allowed them to mark authentic pieces of student work. Their experiences of marking summative assessments while on practicum, however, were quite varied with some gaining knowledge from the experiences of marking students' work, but others not having the opportunity to do marking. However, all teachers said that they wanted more practise and experience marking student work, particularly marking against achievement standards, as they felt this developed their knowledge both of the standards and of how to make judgements:

Well I would really like to do some more marking and then get some feedback on it. That way I can get some experience of what I'm doing right and what I'm not doing right. [Ryan, I3]

All beginner teachers focused primarily on NCEA assessment when they talked about what they knew of marking and making interpretations about student learning. They displayed some knowledge of some of the standards used for NCEA and of the principles of using a standards-based assessment system. This

included some understanding of the structure and format of tasks and associated marking. However, all teachers raised concerns about their lack of knowledge with respect to technical aspects of marking student work against NCEA achievement standards. Teachers found this to be quite a complex process and many struggled with it, doubting their ability to be fair and consistent in applying the levels of achievement described in the standards. They talked about the perceived vagueness of achievement standards which defined performance or proficiency at four levels, and their difficulty in making decisions about student work, particularly differentiating between levels. They all articulated a desire to have more of a focus on marking NCEA tasks in their ITE programmes. For example, at the end of her ITE year, Elise commented,

> But I still would have liked to have more practise with the externals, like marking NCEA. NCEA marking. I'm all right with the rest of the marking, like formative assessment and going back to the stuff they don't know. It's fine. I don't mind that. It's just being very correct in your grading ... I was quite accurate with my grading, but still having that practise and knowing that you're doing it right instead of having to ask people. [Elisa, I4]

Moderation was briefly discussed by teachers when they talked about consistency with marking. However, no beginner teachers in this study articulated a clear understanding of how NCEA moderation worked. They talked about moderation only in terms of its effects on individual teachers, rather than its role in providing robustness and quality assurance for the NCEA system nationwide. Many of the teachers conceptualised moderation as only a process to check up on teachers, and many doubted its effectiveness in ensuring consistency across the country. Negative discussions about moderation between ATs, particularly after they had just received moderation results from NZQA, seemed to have contributed to these beliefs.

By the end of the ITE programme, the teachers in this study could identify particular stances or priorities they applied when marking and making assessment judgements. Some had developed their own practices for dealing with challenging marking and others admitted to not knowing what to do when they encountered work that was not straightforward to mark. For example, all of the beginner teachers talked about their frustrations when they realised their students did not

communicate their ideas clearly enough in an assessment task, yet probably understood more than their written work had suggested. Often this was attributed to a lack of students' test-savviness (which teachers felt they could address), or possibly the students' lack of literacy skills. However, the teachers expressed frustration about how to mark such work. They felt constrained by the evidence from the assessment task.

> There is a lot of grey area where the kid won't word things very well, and it's difficult to know when you're marking it because they get the idea behind it all, by taking them literally or not. I am of the opinion that you should go with more as they understand it … a lot of it comes down to the literacy levels. They may know what they want to say, but they can't express it on paper. [Ryan, I4]

Opinions between teachers varied over whether students should be given the benefit of the doubt in cases where the teacher was confident that students understood the content but did not clearly articulate an acceptable answer, particularly if the teachers felt the students needed encouragement. In some instances teachers talked about "tweaking the marking" to either give the student an inflated grade in order that they feel positive about their learning, or give them a lower grade than they deserved in order to try to motivate them to work harder. More about this is discussed in Section 7.3.2. However, all teachers did concede that for summative assessment tasks that 'counted', that is internally assessed NCEA standards, it was important to only give the students credit for what they have clearly communicated.

In summary, this section has described teachers' knowledge of marking and the interpretation of student work. Overall, the teachers were not confident in their knowledge of interpreting student work or their ability to mark it accurately but did show an understanding of the use of criteria and marking schemes. The following section describes teachers' knowledge of quality in summative assessment.

5.5 Knowledge of quality summative assessment

All beginner teachers made mention of at least one aspect of quality in assessment. This section covers quality in relation to the principles of quality summative assessment (Section 5.5.1), task design (Section 5.5.2) and teacher

judgement (Section 5.5.3). Aspects relating to task design and teacher judgement have been discussed in earlier sections, but in this section the focus is on quality.

5.5.1 Principles of quality summative assessment

There was only limited evidence in this study that beginner teachers considered validity or reliability when summatively assessing students. Teachers did not make much use of the terms validity, reliability or dependability (these terms were used a total of eight times across all of the interviews in the study). However, there was some evidence of them considering whether assessments were of good quality.

The teachers raised issues to do with validity when they discussed the importance of assessment tasks aligning with what had been taught. They all talked about their need to have a good understanding of science knowledge and they linked this to their ability to assess the students' understanding of science content (Section 5.2.1). From this it could be inferred that they had an understanding and knowledge of validity, as they were aware of matching what had been taught with what was assessed. They did raise the difficulty in assessing some of the content and skills they were teaching, which also indicated their concerns over validity in assessment.

Four of the teachers highlighted the needs of individual students. They felt strongly that they should make sure the curriculum level used in any summative assessment task was appropriate for the learners. Ryan discussed one of his experiences of teacher expectations of low ability classes, where

> it was taken for granted that most of them [the students] weren't going to pass the tests, so they [the teachers] didn't worry too much about it. So they just tried to get as many basic concepts into them as they could, so they were less test oriented because they [the students] were in the 'too hard' basket. [Ryan, I4]

Ryan was concerned by this approach because he believed that every student needed to have summative assessment tasks that were appropriate for their cognitive level. Elisa talked about a multi levelled test she had trialled as a way to avoid setting lower ability students up for failure. In the same task she included material from three different science curriculum levels in order to give all students a chance to gain some marks.

170

> Yes, I think it [will] be quite hard because with your junior classes and all the different levels … like with the tests, we tried to rejig it a little [when revising it]. That way keeping it quite similar to how they had it, which is one test at level 3 [NZC level], and then another bit at level 4 to 5. And then they just put it all together and students can just go as far as they can. [Elisa, I2]

There was little commentary about the requirement for reliability in science summative assessment, although some teachers spoke about the importance of using a range of assessment types within any one class as part of acknowledging that some strategies work well for some students but not others (Section 5.3.1). No mention was made of check-marking for junior science assessments.

Overall, the attitude of all the beginner teachers towards validity and reliability in assessment for junior science could be described as *laissez–faire*. Three teachers noted that they were specifically told not to worry too much about the assessment of junior classes when they raised concerns about the assessment tasks that were to be used. However, within the NCEA system, all teachers understood the importance of some sort of rigour, given its high stakes nature. For senior NCEA classes, the NCEA system provides assurance of validity and reliability through its moderation system so the teachers relied on that. As outlined in Section 5.4, all teachers lacked knowledge and confidence with respect to technical aspects of NCEA, marking NCEA tasks, and moderation. They appeared unaware of the real purpose of the NCEA moderation system as a quality assurance mechanism.

However, all teachers seemed attuned to the principles of ethical and fair assessment practice. The impact of their beliefs about this will be further discussed in Section 6.2, but, in essence, all teachers understood the need for all students to have a fair opportunity to give evidence of what they had learned. They raised issues to do with the matching tasks to appropriate curriculum levels (Section 5.2.2) as well as the literacy demands of summative assessment tasks (Section 5.3.2) stating that these could deem the tasks unfair for some students. Fairness and ethical assessment were seen as essential elements for quality assessment.

5.5.2   Task design
As discussed in Section 5.3.1, each of the teachers in this study talked in detail about a range of summative assessment strategies. With respect to quality in

assessment design, however, the beginner teachers appeared to have little understanding of how to design high quality tasks, which is not surprising given their limited understanding of the principles of quality assessment (as outlined in Section 5.5.1). All teachers viewed most of the summative assessment strategies as being equally valid ways to assess science knowledge, understanding and skills. In the interview discussions, most of them made no comparison of the appropriateness of item types or assessment strategies when discussing particular science summative assessments, although Mary and Susie did describe their choices of item types that better suited their junior students, once they were working in their first jobs. Four teachers did talk about tweaking tasks to better match their junior students' interests and strengths, including Wiremu who spoke of changing the language of the assessment (to te reo Māori[1]) in order to better suit Māori students. They all acknowledged that they were developing confidence in designing assessment tasks over time.

The teachers almost exclusively used pre-written tasks for NCEA assessment, with only Susie developing her own tasks. This meant there was little opportunity for teachers to gain a deeper understanding of the task development process as they did not have the opportunity to write NCEA tasks and gain feedback about quality through the school's quality assurance system. In Susie's case she was able to experience and benefit from the quality assurance system in her school.

The samples of summative assessment tasks provided by the teachers over the course of this study confirmed their lack of knowledge about quality, with most tasks being in the form of formal tests which had narrow curriculum focus and low cognitive demand, mostly focusing on knowledge recall. As they talked about the tasks they had designed it appeared that the teachers had very little by way of a quality framework to refer to and were basing their assessments on how they remembered being assessed as students.

5.5.3 Teacher judgement
The beginner teachers only briefly acknowledged the moderation systems set in place to assure quality by check marking and grading for NCEA in their schools, with little demonstrated understanding about how this system operated. They did

---

[1] te reo Māori is the indigenous language of New Zealand, with status as an official language.

show an understanding that quality assessment requires the careful use of criteria (Section 5.4). They all found making judgements and deciding on grades challenging but commented that the experience of working with real student work helped them develop confidence and accuracy. The teachers did also agree that the use of feedback is very important for quality assessment, as it helped students see the justification for the judgements they made.

In summary, teachers were very aware of the need to be fair and ethical in their summative assessment judgements and they generally showed understanding of the use of clear criteria in a marking scheme and making judgements based on interpretation of evidence. Although they rarely used the terms validity or reliability, they did demonstrate developing knowledge of these principles.

5.6 Summary

This chapter has described knowledge related to summative assessment that beginner teachers described during the interviews and demonstrated through artefacts over the time of the study, with the purpose of showing the overall breadth of beginner teacher knowledge about summative assessment. The knowledge that teachers demonstrated of purposes of assessment, what to assess, summative assessment strategies, assessment interpretation, and quality in summative assessment was wide ranging. Not all teachers presented the same knowledge and understandings, but they had shared understandings for many aspects. This chapter allowed a mapping of what was said and presented with key ideas outlined in the literature review. Details of the range and timing of development of this knowledge for individuals is presented in Chapter 6.

Chapter 6: Summative assessment literacy: Development over time

This chapter presents the findings pertaining to the development of summative assessment literacy dimensions for individual beginner teachers over time. Over the period of the study, the beginner teachers were interviewed five times about their knowledge and practices with respect to summative assessment. They completed three questionnaires and were given the opportunity to provide artefacts of their summative assessment practices. All of this data was combined and analysed using the SALRubric (full details of the SALRubric are provided in Section 4.4.2) in order to document shifts and development in dimensions of the teachers' summative assessment literacy. The sections covered in this chapter are trends in the development of individual dimensions of summative assessment literacy (Section 6.1), development for individual teachers, including three examples (Section 6.2), and contributors to beginner teacher development over time (Section 6.3).

6.1 Trends in the development of individual dimensions

In this section trends for teachers' SALRubric scores for different dimensions of assessment literacy over time are reported. Scores range from 1 (Novice), 2 (Advanced Beginner), 3 (Competent), 4 (Proficient), to 5 (Expert). Shifts for individual aspects of assessment literacy are summarised in Tables 7–17 below, and are grouped into three categories: knowledge of assessment (the first five dimensions of the SALRubric) (Section 6.1.1), understanding the context for assessment (dimensions six, seven, and eight of the SALRubric) (Section 6.1.2), and recognising the impact of assessment (the last two dimensions of the SALRubric) (Section 6.1.3).

6.1.1 Knowledge of assessment

Beginner teachers commenced their initial teacher education programme with some prior knowledge of summative assessment gained through their own educational and life experiences. Through the course of their ITE study and practicum experiences they gained further knowledge and experience pertaining to assessment. The following section describes shifts in assessment literacy dimensions that relate specifically to the knowledge teachers possessed, and upon which they acted. This relates to the first five dimensions of the SALRubric.

6.1.1.1 Ability to describe assessment

Each beginner teacher's ability to describe assessment was of interest because this gives an indication of their overall view of assessment including its complexities. This view of assessment may in turn determine the teacher's focus and approach to assessment (Pajares, 1992, Stuart & Thurlow, 2000). Data for this dimension is in Table 7 below.

Table 7: SALRubric Scores for Dimension 1: Ability to Describe Assessment

Beginner teacher	Interview 1		Interview 2	Interview 3		Interview 4		Interview 5
Bella	1		2	2		2		3
Benjamin	1		2	2		2		2
Elisa	1	Practicum	2	2	Practicum	2	Teaching experience	3
Kate	1		2	2		3		3
Mary	1		2	2		3		3
Susie	1		2	2		3		3
Ryan	1		2	2		2		2
Wiremu	1		3	3		3		3

At the beginning of the study, all beginner teachers in this study provided evidence judged to be a Novice for their ability to describe assessment and being able to describe assessment in very broad terms based on their own experience. Many of the teachers struggled in the first interview to articulate a definition and most referred to tests:

> I guess it is a judge of someone's ability in a certain area, maybe not so much their ability but … It's hard to explain … Their ability to transform what's in their head onto paper. [Benjamin, I1]

All eight beginner teachers showed a shift between Interview 1 and Interview 2 from Novice to Advanced Beginner, with the exception of Wiremu who was judged to be Competent. At the Interview 2 time point all teachers in the study were able to describe summative assessment, formative assessment and diagnostic assessment using simple language. Wiremu defined it in a more nuanced way, including an understanding of ethical components. In Interview 4, for example, Wiremu produced artefacts of summative assessment tasks he had designed, and he talked about the

ethical decisions he had to make in their design. He also described the purposes and effects of assessment:

> My understanding of the assessment would be, first of all to break it up into diagnostic, formative and summative assessment. Are you figuring out what they know? Are you figuring out, trying to give them a test in order to give them more feedback and feed forward? Or is it just a pure assessment to gauge where they are, and to compare them to their peers? There needs to be some way of measuring and I'm not a big fan of excessive summative assessment, but there does need to be some for every unit you do. There has to be something where you can show — this is their work — and you can relate it to the quality of their work. Because essentially you need some kind of evidence for learning … and summative assessment doesn't necessarily show all of the learning that has taken place, but it is something, and the systems are already in place to get that information. [Wiremu, I4]

Three teachers showed further measureable shift later in the year, with a further two showing shift after their first six months teaching (at Interview 5).

By the end of the study, six of the eight teachers were able to describe assessment types and assessment effects and uses in a more integrated way, including demonstrating an understanding that ethical decision making is required. The other two teachers were able to describe summative assessment, formative assessment, diagnostic assessment and their purposes, using simple descriptions but did not demonstrate an integrated understanding of the purposes, effects and uses of assessment with students.

6.1.1.2   *Knowledge of purposes of summative assessment*

The knowledge beginner teachers held regarding the purposes of summative assessment was of interest as this knowledge was likely to determine the purposes for which assessment was used in their classes, potentially enhancing or constraining the uses the teachers made of assessment. The breadth of beginner teacher's knowledge about the purposes of summative assessment was described in Section 5.1. Table 8 summarises the individual teachers' development in this dimension over time.

Table 8: SALRubric Scores for Dimension 2: Knowledge of Purposes of Summative Assessment

Beginner teacher	Interview 1		Interview 2	Interview 3		Interview 4		Interview 5
Bella	2		3	3		3		3
Benjamin	1		2	2		3		3
Elisa	1	Practicum	2	3	Practicum	3	Teaching experience	3
Kate	2		2	2		3		3
Mary	1		2	2		3		3
Susie	1		3	3		3		3
Ryan	1		1	2		2		3
Wiremu	2		3	3		3		3

In the first interview in this study, five beginner teachers gave evidence of being able to articulate the purposes of assessment using general and vague terms. For example, Elisa struggled to describe the purposes in response to a direct question:

> Assessment is kind of looking at what you could do in regards like what you're expected to be able to do … I don't know how to explain it, it's hard. [Elisa, I1]

In Interview 1 three of the teachers were able to clearly articulate a number of specific purposes for summative assessment such as informing pedagogical decision making, reporting, qualifications, and school uses such as streaming. Over time the beginner teachers developed a deeper knowledge of purposes of assessments relatively quickly with six of the teachers demonstrating shift between Interviews 1 and 2. Only two teachers showed a shift between Interviews 2 and 3, but by the end of the ITE year, seven out of eight teachers were scored Competent, that is, they considered a range of assessment purposes and understood that an assessment task can be used for a range of purposes. The eighth teacher demonstrated evidence of Competent understanding after teaching for six months. Below is an example of a quote showing Elisa's description of some of the purposes of summative assessment at the end of her ITE year:

> I've learnt it's in a lot more contexts than I originally thought … So it's basically the process we find out what the students actually know … So just being able to alter your teaching to direct what the

content is to what their needs are, basically. The most dominant [purpose] would be report writing … for the Year tens, to see how they're doing and what needs to be addressed, and how much they have learnt to go on to NCEA. And then for the NCEA [students], the assessment was more preparing them for the exams ... to see what they know, and what they don't know, misconceptions … I think [my views on the purposes of assessment] have changed a lot, because I now see it as a more useful tool rather than something they just have to do at the end of the topic, just to follow the report or something. [Elisa, I4]

Elisa had shifted in her thinking over the course of the year, and her understanding of the purposes of summative assessment became broader, as she acknowledged it could serve a range of purposes, rather just something that happens at the end of a topic.

By the end of the study (18 months after the beginner teachers commenced their ITE course) all beginner teachers considered a range of specific purposes for summative assessment, including informing pedagogical decision making, reporting progress, qualifications and school uses. They provided evidence of understanding that the same assessment task could serve a range of purposes.

6.1.1.3 *Knowledge of what to assess*

Knowledge of what to assess is of interest because teacher understanding of the links between curriculum, subject content, standards and assessment is critical for validity in assessment. Because of the responsibility for assessment decision making given to classroom teachers, it is important that they understand these links. Data for teacher development in this dimension is provided in Table 9.

Table 9: SALRubric Scores for Dimension 3: Knowledge of What to Assess

Beginner teacher	Interview 1		Interview 2	Interview 3		Interview 4		Interview 5
Bella	1		3	3		3		3
Benjamin	1		2	2		2		3
Elisa	1	Practicum	2	2	Practicum	3	Teaching experience	3
Kate	1		2	2		3		4
Mary	1		2	2		2		3
Susie	2		2	2		3		3
Ryan	1		1	1		2		2
Wiremu	1		2	2		3		3

Early in the study, beginner teachers provided little evidence of the knowledge of what was important to summatively assess. All teachers except one were judged Novice at Interview 1. These teachers considered summative assessment to be a generic activity and showed no indication that assessment should be linked to the specifics of what was taught. Setting some sort of test at the end of a topic was the expectation by these teachers, and in Interview 1 they were not yet linking the content of tests to the learning objectives, derived from the curriculum. A noticeable shift occurred early in their ITE as was documented at Interview 2. By Interview 2, six of the eight teachers showed a shift, with most being able to talk about assessment criteria that linked to the learning objectives and the science standards. No beginner teachers showed significant knowledge development between Interviews 2 and 3. In Interview 4 a number of the teachers demonstrated development in this area, with five teachers demonstrating a shift after their second practicum. The two interviews where shift was documented in the majority of teachers both followed practicum experience, during which they had been exposed to summative assessment in the classroom. Further shift was evident in three teachers after six months of teaching. By the end of the study seven of the teachers showed that they identified specific science knowledge and skills that were important to assess, and were comfortable in critiquing summative assessments for this content. They considered the importance of linking assessment to standards criteria and to what was taught. At the end of the study, Kate provided documentary evidence that indicated she was operating as Proficient, that is, she considered the need to identify specific science skills and knowledge that are important to assess, could see need for alignment between NZC, teaching and assessment, considered validity and could accurately critique for content within assessment tasks, including NCEA tasks.

6.1.1.4   *Knowledge of assessment strategies and task design*
Knowledge of summative assessment strategies and task design is of interest because beginner teachers are expected to be able to make decisions about the strategies they will use to assess students, and are often responsible for the design of assessment tasks. Therefore knowledge of assessment strategies and task design is likely to affect the quality of their assessment practice. Data for teacher development in this dimension is presented in Table 10 below.

Table 10: SALRubric Scores for Dimension 4: Knowledge of Assessment Strategies/ Task Design

Beginner teacher	Interview 1		Interview 2	Interview 3		Interview 4		Interview 5
Bella	1		2	2		2		2
Benjamin	1		2	2		2		3
Elisa	1	Practicum	2	2	Practicum	2	Teaching experience	3
Kate	1		2	2		3		4
Mary	1		2	2		2		3
Susie	1		3	3		4		4
Ryan	1		1	1		2		2
Wiremu	0		3	3		3		3

Early in this study, all beginner teachers showed limited knowledge of assessment strategies and the design of assessment tasks, only referring to generic forms of summative assessment such as tests and examinations, as well as 'internals'. For example, Susie commented,

> The assessment type is very specific depending … And so in that sense you need a wide range of understanding of assessment so that you can then make it the most effective way of learning. [Susie, I1]

Although Susie said teachers need a wide range understanding of assessment she only mentioned tests and examinations as summative assessment types.

However, considerable shift occurred early in their ITE year with six out of eight beginner teachers showing a shift between Interviews 1 and 2, moving to Advanced Beginner, that is, considering the use of a range of assessment types, but still not confident to design these themselves. Five of the teachers were still Advanced Beginner at the end of their ITE year, indicating that when they started their first teaching positions the majority of teachers were not confident in designing summative assessment tasks. However, there was evidence of further shift after the teachers had been teaching in their schools for six months, so that by the end of the study, six teachers appeared confident and competent in using a range of strategies and in the design of assessment tasks. Furthermore, two teachers were judged as Proficient, as they provided evidence of considering the design of summative

assessment tasks that best match SLOs/ standards and students, using authentic scientific practices, and she understood the need for assessment tasks to be critiqued for validity and reliability. Susie provided an example of an NCEA task she had developed, and talked about the parameters of such a task:

> I designed the year 13 Biotech one, but you are still kind of set within guidelines, you had to do a report, you had to research. [Susie, I4]

At this interview [I4] Susie also described a range of alternative assessment strategies and the quality assurance processes she followed, hence demonstrating considerable development.

Beginner teachers' confidence levels in this dimension appeared to be related to the level of autonomy schools extended to their beginner teachers. Those teachers in schools that exerted a high degree of control on teachers' summative assessment practice were less confident, whereas teachers who were given more autonomy certainly appeared more confident. All beginner teachers talked about how experience was helpful for their development of sustained confidence.

It was surprising to see teachers reaching Proficient category for this dimension, and they attributed this to particularly helpful ATs who involved them fully with summative assessment for NCEA in their classes during practicum experiences, and to first employing schools in which young teachers were encouraged to make their own decisions and were also provided with the support they needed.

6.1.1.5   *Knowledge of assessment interpretation*
This dimension is of interest as an essential element of summative assessment. It encompasses marking and grading, and the interpretation of what the results mean both for the student and for the teacher. Science teachers in New Zealand tend to use summative assessment tasks that require students to provide description and discussion of concepts rather than simply testing their memorisation of facts, so marking these tasks requires accurate teacher judgement and interpretation, as the answers are not necessarily simply right or wrong. Interpretation occurs at an individual level for student work, and it is important for teachers to interrogate learning progress as well as grades for individual students. Interpretation also occurs at class/school level as summative assessment results are conflated and this data is analysed for class levels within science departments. Table 11 below illustrates the development of teachers in this dimension.

Table 11: SALRubric Scores for Dimension 5: Knowledge of Assessment
        Interpretation

Beginner teacher	Interview 1		Interview 2	Interview 3		Interview 4		Interview 5
Bella	1		3	3		3		3
Benjamin	2		2	2		3		3
Elisa	1	Practicum	2	2	Practicum	2	Teaching experience	3
Kate	1		2	2		3		3
Mary	1		1	2		2		3
Susie	2		2	2		3		3
Ryan	2		2	2		2		2
Wiremu	1		2	2		3		3

At the beginning of the study, five beginner teachers were scored as Novice for this
dimension, as they considered that a marking scheme was all that was required to
mark student work and that marking was straightforward, requiring little thought.
However, three beginner teachers were scored at Advanced Beginner as they were
aware of their role as teachers in making judgements about student work. Over time
there was development in teachers' knowledge of the importance of their roles in
fuller interpretations of student work. Four beginner teachers showed a shift in their
knowledge in Interview 2, but only one showed further shift in Interview 3.

For example, by Interview 3, Mary had shifted from very little understanding of how
to mark and make judgements about student work, to understanding the need for
consistency in using a marking scheme accurately:

> And it's cool because on the marking schedule they've got keywords
> that they have to have, and a little sentence about it. So you kind of
> know what you are looking for … But you can tell from the
> assessment schedule what grade they are at by what they include.
> Okay you can make a grading schedule and all teachers could have the
> same one, but actually moderating between the grade I give somebody
> and the grade that somebody else does. [Mary, I3]

Four teachers made further shift by Interview 4, and by the end of the study all but
one teacher were scored at Competent, as they considered the importance of using
marking schemes accurately and making judgements based on the interpretation of
evidence. These beginner teachers showed that they had grasped the idea that

marking was not black and white, but that their interpretation of evidence was central to the process. One beginner teacher entered the programme as an Advanced beginner but provided no evidence of measureable shift throughout the study.

In this section, which focused on the development of teachers' knowledge of assessment over a number of dimensions, it is significant to note that for all teachers there was significant learning of new knowledge, and a pattern of more gains being made at certain times in the ITE programme. Evidence of increased knowledge was present after teachers had completed a practicum experience, with less shift occurring between Interviews 2 and 3 during which time they were on campus in university classes and had no classroom exposure. A number of teachers, though not all, showed shifts in some aspects of assessment knowledge once they were teaching in schools.

6.1.2 Understanding the context for assessment

The following section describes the shifts in assessment literacy dimensions that have strong links to the context of assessment in New Zealand schools. Three dimensions are described: NCEA assessment, preparing students for summative assessment, and using summative assessment formatively.

6.1.2.1 NCEA assessment

NCEA is the secondary qualification achieved by New Zealand school leavers as described in Section 2.2.2. Classroom teachers are responsible for much of the decision making and implementation of NCEA, a high stakes summative assessment. There is therefore considerable interest in teacher development with respect to NCEA assessment. Data for teachers' development in this dimension is given in Table 12.

Table 12: SALRubric Scores for Dimension 6: NCEA Assessment

Beginner teacher	Interview 1		Interview 2	Interview 3		Interview 4		Interview 5
Bella	1		2	2		2		3
Benjamin	1		2	2		2		2
Elisa	1		2	2		3		3
Kate	1	Practicum	2	2	Practicum	3	Teaching experience	3
Mary	1		2	2		2		3
Susie	1		2	2		3		4
Ryan	1		1	1		2		2
Wiremu	0		2	2		3		3

Early in the study, beginner teacher knowledge of NCEA and their role in assessing achievement standards was very limited for all teachers, including those who had themselves been assessed for NCEA as secondary students. Seven of the teachers were scored as Novice and one provided no evidence of understanding the basics of NCEA. For example, in the first interview Kate talked about her knowledge of NCEA based on her experience of the qualification as a student, but did not appear to have an understanding of the teacher's role in the NCEA system:

> I think me being a bit younger and going through the NCEA, the standards levels 1, 2 and 3, I think it's quite familiar to me. [Kate, I1]

However, at the end of her ITE year, Kate was able to describe in detail the full process of assessment against an achievement standard for NCEA, including adjusting tasks to suit her class, running the assessment, marking student work, and having marking moderated. This was because on her second practicum she was involved in the assessment of a standard with her Year 12 class, and she experienced each step of the process.

Beginner teachers in this study showed shift in their knowledge of NCEA, particularly after each practicum experience (as evidenced in Interviews 2 and 4), with five of the eight teachers showing shift in Interview 2 and five showing shift in Interview 4. Most teachers were at least Competent after teaching in a school for six months, that is, they considered the need for alignment between teaching, NCEA standards and assessment tasks, and showed an understanding of the marking and moderation of NCEA tasks. These teachers adjusted or adapted assessment tasks to suit their students. Two teachers were scored as Advanced Beginners at the end of the study. These two teachers were both working in schools where their summative assessment practice was dictated to them by senior teachers. Coincidentally, both of these teachers had experienced at least one practicum in a similarly regimented school where they had been given little or no experience with NCEA assessment. They had not yet experienced any autonomy in summative assessment practice.

6.1.2.2 *Preparing students for standards-based assessment*
For good summative assessment practice, teachers need to educate their students about the nuances of the NCEA system as well as prepare them for the summative assessments they will face. This dimension is of interest because of the students'

dependence on their teachers for this preparation. Table 13 below shows the development of beginner teachers in this dimension.

Table 13: SALRubric Scores for Dimension 7: Preparing Students For Standards-based Assessment

Beginner teacher	Interview 1	Practicum	Interview 2	Interview 3	Practicum	Interview 4	Teaching experience	Interview 5
Bella	1		1	1		2		3
Benjamin	1		1	1		1		2
Elisa	1		2	2		3		3
Kate	1		2	2		3		3
Mary	1		2	2		3		3
Susie	1		3	3		3		3
Ryan	1		1	2		2		3
Wiremu	1		2	2		3		3

All beginner teachers were scored as Novice at the beginning of this study as they did not provide evidence of thinking about students' assessment literacy and did not consider that teachers played a role in preparing students for assessment. Ryan, for example, was vocally opposed to teachers preparing students for assessment, as he deemed this to be "teaching to the test". He commented on the practice he observed while on his first practicum:

> Teaching and catering just for the exam was blatant and it was ridiculous. [Ryan, I2]

In his final interview, however, he explained how he did help students prepare for standards-based assessment by helping students understand criteria and related marking schedules in the tutorials he ran at lunchtimes.

It was clear that practicum experiences made a difference for a number of teachers for this dimension, as five out of eight teachers showed a shift between Interviews 1 and 2, and five out of eight teachers showed a shift between Interviews 3 and 4. The shifts at other data collection points were fewer with only one teacher showing a shift between Interviews 2 and 3, and three teachers showing a shift between Interviews 4 and 5. By the end of the study seven of the eight teachers considered the importance of formative practices in preparing students for summative assessments and

understood they had a role to play in this. One teacher did consider the need of students to understand assessment processes and procedures, including the criteria to which they were being measured, but was yet to explore his role in this.

In this study all teachers were strongly opposed to teaching to the test, which many of them felt was the way some teachers prepared students for assessments. They all voiced the opinion that this practice undermined the NCEA summative assessment system.

6.1.2.3   *Using summative assessment formatively*

Because of the commitment to formative assessment in New Zealand education policy, this dimension is of interest. The New Zealand curriculum clearly prioritises the use of assessment to improve students' learning and teachers' teaching; therefore teacher understanding and using summative assessment information to inform their decisions is relevant in the development of assessment literacy. Teachers' development in this dimension is given in Table 14 below.

Table 14: SALRubric Scores for Dimension 8: Using Summative Assessment
          Formatively

Beginner teacher	Interview 1		Interview 2	Interview 3		Interview 4		Interview 5
Bella	1		3	3		3		3
Benjamin	2		3	3		3		3
Elisa	2	Practicum	3	3	Practicum	3	Teaching experience	3
Kate	2		3	3		3		3
Mary	1		2	2		2		3
Susie	2		3	3		3		3
Ryan	2		2	2		2		2
Wiremu	2		3	3		3		3

Most teachers in this study considered summative assessment as affording opportunities to learn about their own teaching and their students' learning, and at the beginning of this study the majority of teachers appeared to be motivated to make use of summative assessment in order to improve their teaching. For example, in the first interview, Benjamin demonstrated that he considered he could use summative assessment formatively when he commented,

> [Summative assessment is] to gain an understanding of where they're
> at and how well the students are doing and the teachers are doing, and
> I guess from there you can make improvements. I guess that would be
> the overarching thing, so where to make changes to improve for the
> teacher, and how to make things improve for the students. [Benjamin,
> I1]

However, at this stage (I1) he was vague about how he would do this. Later in the
study, Benjamin showed further development as he identified how he could do
this more specifically:

> I think, for us, basically they are to guide us as teachers. The students
> might be having some misunderstandings of concepts, or struggling,
> and then we can teach to that area. And then also I guess you have to
> have assessment to see how well the students are doing. Well, that's
> why … for qualifications, for next year and all the rest of it. But I
> think the main purpose … to see where the main gaps are, where our
> teaching could improve, or where the students just need more
> learning. [Benjamin, I4]

Considerable shift was seen in their conceptions in this dimension after the first
practicum experience, with seven beginner teachers showing a shift between
Interviews 1 and 2, but only one more shift in this dimension noted for the rest of the
study. By the end of the study, seven of the teachers considered summative
assessment as an important source of information in both providing feedback on their
teaching and feed forward for decisions on further teaching. One teacher, Ryan, did
understand that he could learn from the summative assessment results of his students,
but did not provide evidence that he was using this to inform his ongoing teaching
decisions.

In summary, this section focused on the context for assessment, and beginner teachers
provided evidence of development in dimensions describing NCEA assessment,
preparing students for standards-based assessment, and using summative assessment
formatively within the New Zealand context. As in the previous section, the impact of
practicum was noted as more shift was evident for the teachers after these practicum
experiences.

6.1.3   Recognising the impact of assessment

This section is concerned with the human side of summative assessment, as teachers
acknowledged that their practices created impacts on their students. Beginner teachers
highlighted teaching as a relational practice, and described their need to be responsive

and show care and respect to their students and others, conducting themselves in an ethical manner. They considered the effects of assessment for students in two dimensions: understanding assessment consequences, and fairness.

6.1.3.1 Understanding assessment consequences

A strong theme evident in the data collected was beginner teachers' understanding of the consequences of assessment for the student. Understanding assessment consequences is of interest because teachers' conceptions of the consequences of assessment is likely to affect the decisions they make about summative assessment. Table 15 below illustrates the development of beginner teachers in this dimension.

Table 15: SALRubric Scores for Dimension 9: Understanding Assessment Consequences

Beginner teacher	Interview 1		Interview 2	Interview 3		Interview 4		Interview 5
Bella	2		2	2		3		3
Benjamin	2		2	2		2		2
Elisa	1	Practicum	2	2	Practicum	3	Teaching experience	3
Kate	1		1	2		2		2
Mary	1		1	2		2		2
Susie	1		1	2		2		2
Ryan	2		2	2		2		3
Wiremu	2		2	2		2		3

Early in this study, four beginner teachers showed limited acknowledgement of the range of consequences of summative assessment for students and teachers, and four were scored at Advanced Beginner stage. Wiremu, for example, did understand that assessment had an effect on students, as he had experienced some traumatic assessment in his own childhood. Over the course of their ITE year, three teachers made no observable shift in their conceptions in this dimension. When compared to other dimensions, the development in this dimension seemed to take more time, and there was not a significant shift for teachers over the practicum period. However, the teachers who shifted in their thinking seemed to do so later in the programme, with only one teacher showing significant shift in the first half year of the study. At the end the study, four of the beginner teachers were scored as Competent, that is, they reflected on the effects of assessment on students, and used assessments to students'

advantage, planning in ways that mitigated harm. These teachers also considered the effects of summative assessment results on teachers and schools. For example, Wiremu was thoughtful about his assessment practice and the potential effects it would have on students:

> Just do small assessment things but try to think about everything involved with it. You know, leading up to the assessment, the doing of it, the learning of things, and then how the assessment itself affects the class, following on from the assessment. [Wiremu, I3]

Wiremu discussed the way he adjusted assessment for Māori speakers (see Section 5.3.2) in attempt to mitigate harm they may experience through being assessed through the medium of English.

The other four beginner teachers did consider the effects of assessment on students (e.g., emotional, motivational) or teachers but at this stage were not using this to inform their planning or decision making.

6.1.3.2   *Fairness*

Fairness is of interest in this study as fair assessment is of importance to students and the wider community. Fairness is a complex construct. Table 16 below shows the development beginner teachers made in this dimension over the course of the study.

Table 16: SALRubric Scores for Dimension 10: Fairness

Beginner teacher	Interview 1		Interview 2	Interview 3		Interview 4		Interview 5
Bella	1		2	2		2		2
Benjamin	1		2	2		2		2
Elisa	1		2	2		3		3
Kate	1	Practicum	1	1	Practicum	2	Teaching experience	2
Mary	1		1	1		2		2
Susie	1		2	2		2		2
Ryan	1		1	1		2		2
Wiremu	2		2	2		3		3

At the beginning of the study seven of the eight beginner teachers provided little evidence that they considered the need for fairness in summative assessment, but all teachers made considerable shifts over time. Shifts in the teachers' considerations of

fairness during ITE was only evident after each practicum, with four teachers making a shift to Advanced Beginner after the first practicum and five teachers making a shift after the second practicum. Three teachers showed further shift after they had been teaching for six months. By the end of the study, five out of eight beginner teachers were judged to be Competent, that is, they considered the need to identify ways that would enable individual students to show their learning, and were willing to investigate different modes by which to assess students against the same specific learning outcomes or standards. However, these teachers still had a limited range of strategies to use. For example, in the fourth interview, Benjamin explained how it could be fairer for students to choose assessment types that suit them:

> Fairness, I guess, it's hard to define fairness because it's the same for everyone, but whether that is fair or not ... So I think that summative assessment is generally [fair] because it is the same for everyone, but maybe a better assessment might be to let students choose whether they want to give a written essay or give an oral presentation. I think that would be more fair. [Benjamin, I4]

The three teachers at Advanced Beginner felt that fairness meant it was important to treat all students exactly the same.

In this section focused on recognising the impact of assessment, the data shows gradual development for all beginner teachers in this study. Teacher shift for both of the dimensions in this section appeared to take more time than other dimensions.

Overall use of the SALRubric has allowed the development of teachers' summative assessment literacy in a number of dimensions to be documented over time. Development was seen to occur for all dimensions, although the development in recognising the impact of assessment was seen to take more time. Most shift was seen after beginner teachers had spent time in schools.

6.2   Development for individual teachers

This section focuses on the development of dimensions of summative assessment literacy for individual teachers, as documented through the use of the SALRubric (Section 6.2.1). Three examples of development are presented (Section 6.2.2), followed by a summary of the patterns for individual development (Section 6.2.3).

6.2.1 Individual teacher development

All data related to each beginner teacher was analysed using the SALRubric. A table summarising the data by teacher is available in Appendix 11. Figure 4 below summarises the development of individual teachers, where dimension scores are displayed in radar graphs. The concentric lines in the radar graphs represent data gathering points: Interviews 1-5. On the graphs Interview 1 is the innermost ring, Interview 2 the next ring out from the centre, and so on to Interview 5, the outermost ring.

Figure 4.   Radar charts of summative assessment literacy development for eight beginner teachers.

Key: 1. Ability to describe assessment   2. Knowledge of purposes of summative assessment   3. Knowledge of what to assess
4. Knowledge of strategies and design   5. Knowledge of assessment interpretation 6. NCEA assessment 7. Preparing students for standards-based assessment   8. Using summative assessment formatively   9. Understanding assessment consequences   10. Fairness
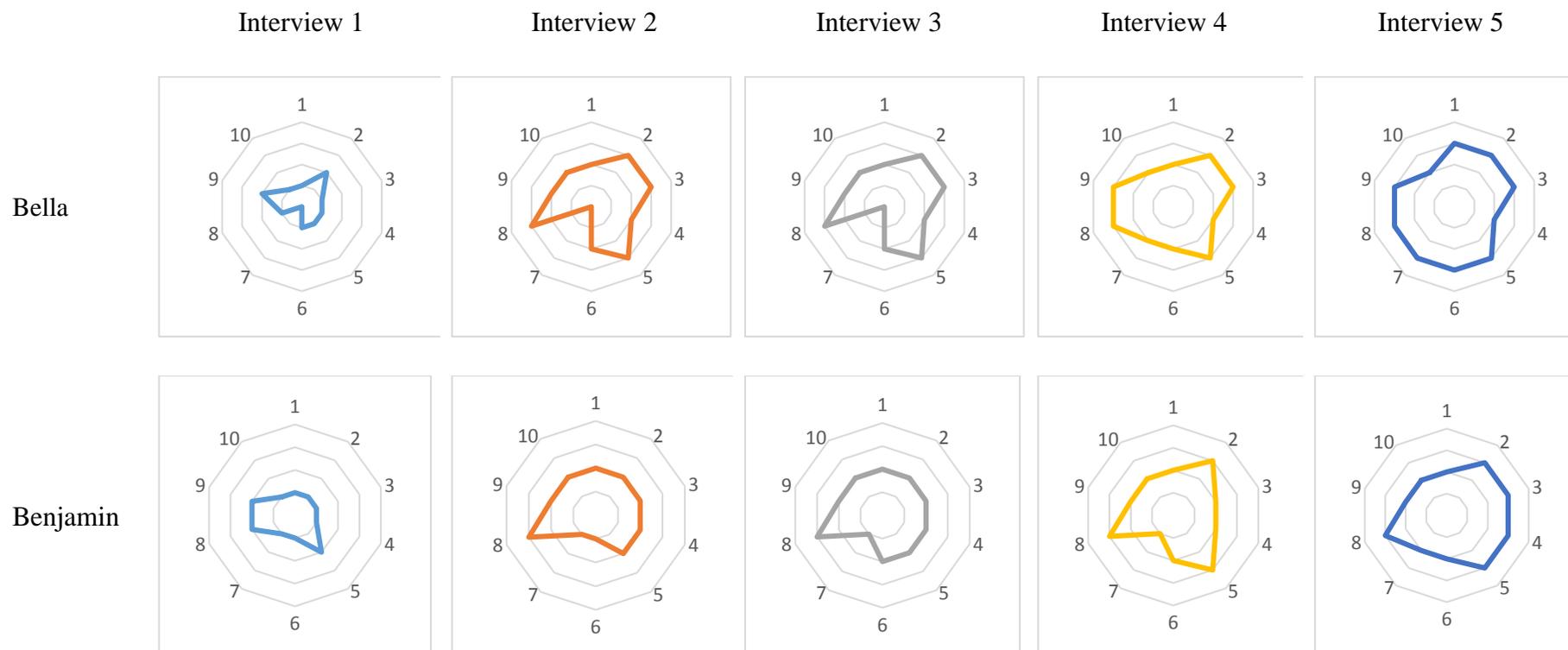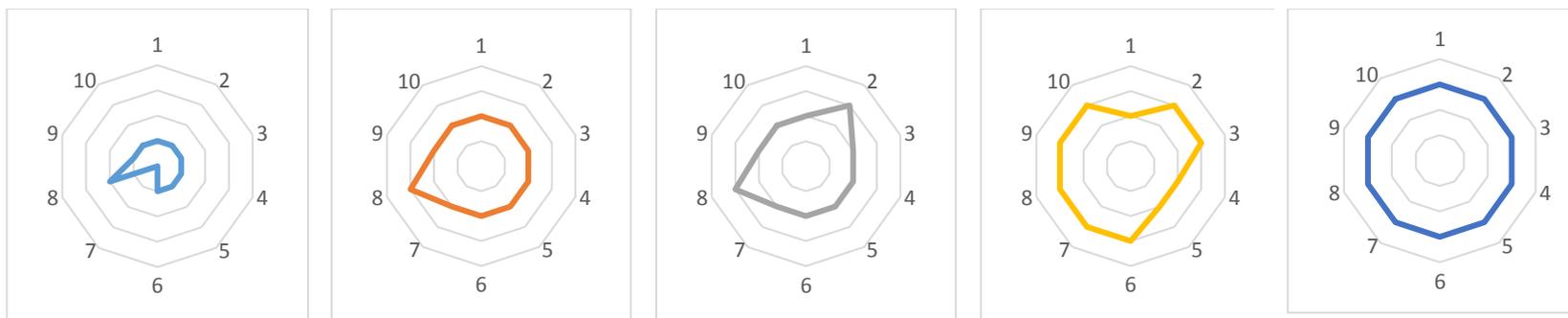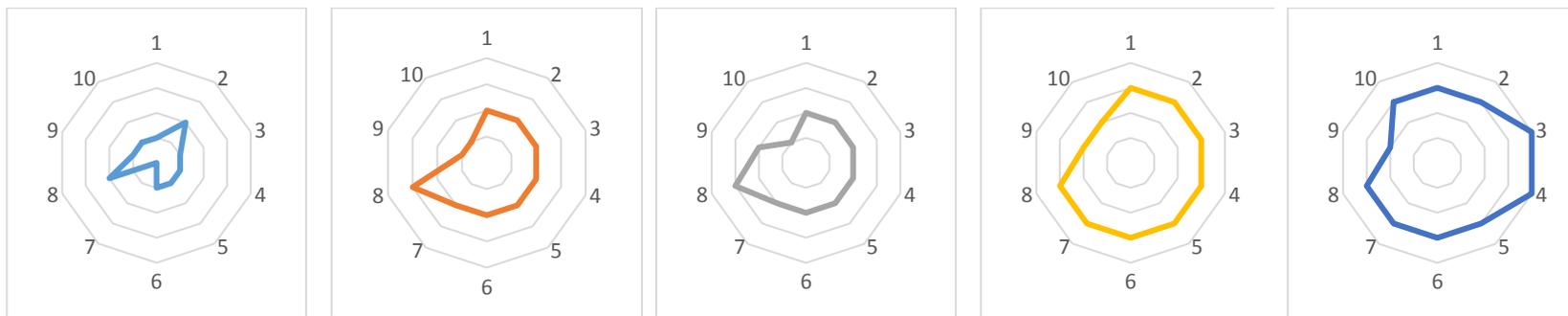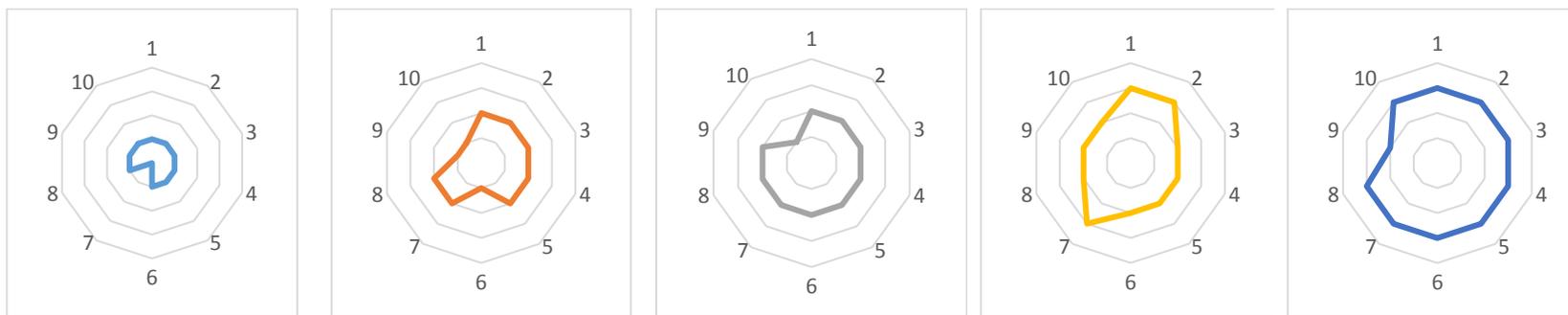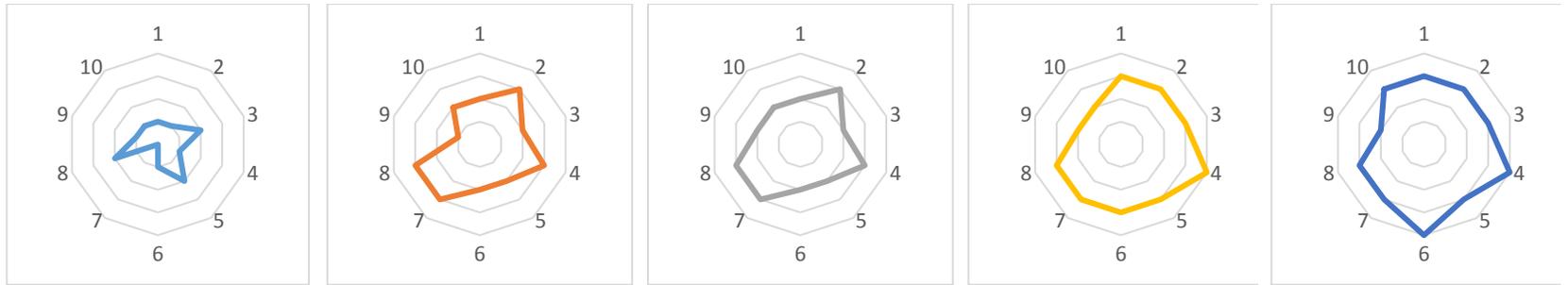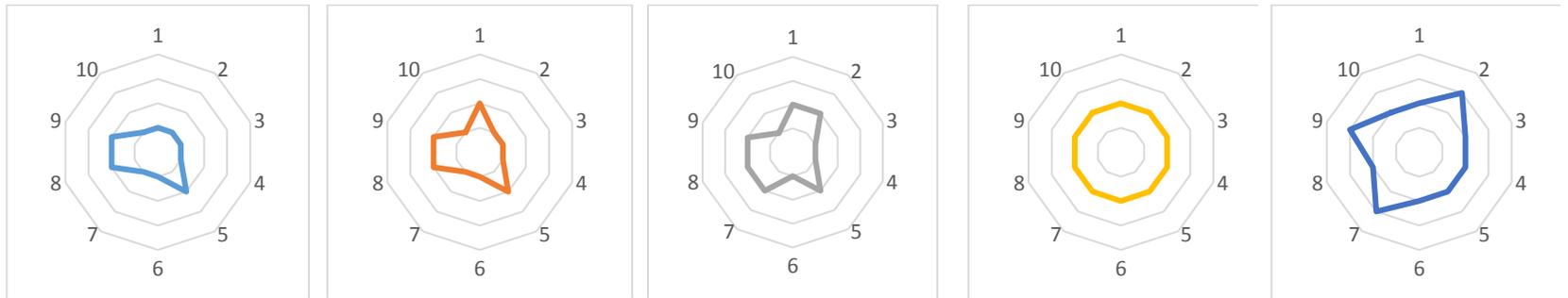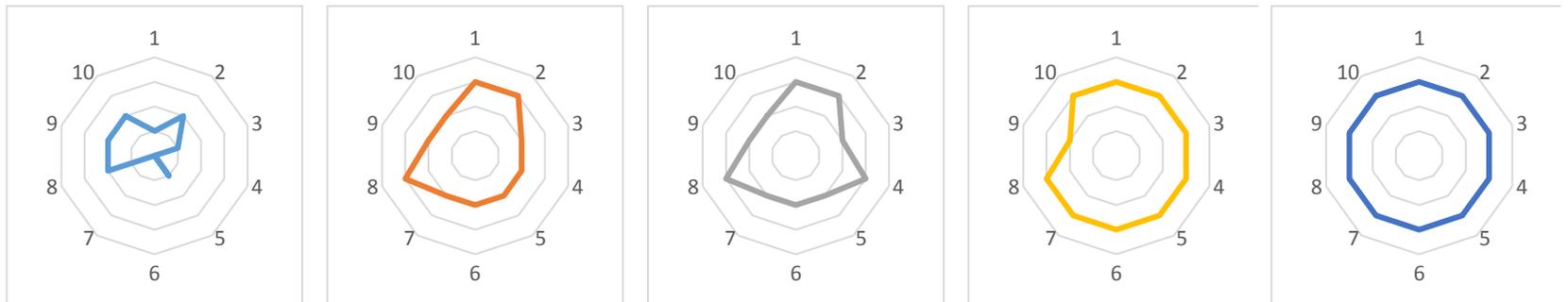
6.2.2 Vignettes of development: Bella, Susie, Ryan

The patterns of development in the 10 dimensions of assessment literacy as detailed in the SALRubric were idiosyncratic across the group although all beginner teachers were more likely to show improvement in their assessment literacy after their practicum experiences. In this section a brief description of the development of three teachers is given. These three teachers were chosen in order to illustrate the variation in the development of beginner teachers' summative assessment literacy. Bella entered the programme least confident of the three and showed significant development early in the study, followed by little change over the next year. Susie was confident and noticeable shift was evident after each practicum experience, and after teaching in her first position. Ryan entered the programme with little confidence and with a negative view of summative assessment. His progress was much slower than that of all other beginner teachers in the study.

6.2.2.1 Bella

Bella entered the ITE programme following her university study immediately after she gained a Bachelor of Science degree with a microbiology major. Both of her parents and her sisters were teachers, so she considered she understood the pressures of her chosen career. Bella's own experience of learning science at school was not all positive:

> My experience in science as a pupil wasn't amazing. I had one good teacher in Year 11, but past that [learning science] kind of lost the spark, you know? So I'm so glad I carried on because I really got enthused about it at uni, and thought that there were better ways to teach it. [Bella, I1]

Bella's orientation to teaching was relational and she talked about prioritising building good relationships with students. She believed that by interacting with them in positive ways she would encourage their enjoyment of science and hence their motivation to learn. She believed that motivation was required if students were to learn, so she continually looked for ways to help motivate students and to provide them with 'fun' activities, learning games, and online interactive programmes.

She was constructivist in her general approach to learning and teaching. That is, she believed that learners built on what they knew and so she planned for learning

and assessment in ways that allowed her to elicit their prior knowledge and respond to this by scaffolding and relating new knowledge to what students already knew and could do. Bella believed her role as a teacher was to fully understand science content and present it in ways that made it accessible for all students, including those with low levels of literacy or numeracy, suggesting that she was also concerned with equity in the science classroom.

Bella maintained that students learnt from each other while engaged in group learning activities, and she explained that she encouraged class discussion and student interactions through group work. This was often achieved through practical and hands-on activities, as illustrated in artefacts she presented. Over the interviews she brought copies of unit plans, lesson plans, and summative assessment tasks she had designed and/or used to illustrate her teaching and summative assessment practice. Her unit plans were interesting and varied, and by the end of her ITE year her documentation for assignments showed that she was starting to think creatively for summative assessment. For example she produced unit titled "Crime Scene Investigation: Who killed Mr X" in which the students learnt aspects of forensic science: separating mixtures, DNA structure, fingerprints, blood types, cell types and structure, and the discussion of a controversial issue. Her summative assessment involved pairs of "detectives" using Go Animate (software which generates cartoons) to explain to the class how they would go about investigating a case they were given. Although the science content assessed in her final summative assessment was limited, the task certainly had potential. This task linked with Bella's beliefs about science needing to be 'fun'. Overall, Bella's commentary suggested that she valued the use of formative assessment strategies during group activities.

Bella maintained that university course work was not very useful, and she needed to learn by trying things out in a practical environment:

> Nothing really stuck…. Even if I have already 'learnt' it I probably won't remember it until I'm in the classroom. I kind of have to practise something to know it. [Bella, I1]

> We had a couple of lectures [about summative assessment] this year before I went [on practicum], but I sort of learnt about it on-the-job, and then saw the importance myself. [Bella, I2]

I got better during prac. I guess [Lecturer Y] has given us more
techniques … but it's not until we use them [that we learn] … I feel
that uni almost slows you down, when in comparison to being on prac.
Because I felt that I learnt heaps on prac. [Bella, I3]

Bella had two practicum experiences in schools where she was given less
experience with summative assessment than she wanted. She was able to observe
a number of summative assessment events in her first practicum, and gained
limited experience marking examinations in her second practicum school. She
was then appointed to her first teaching position in the school where she had her
first practicum. This school mostly prescribed common summative assessment
tasks for her classes.

Table 17 below summarises Bella's development in dimensions of summative
assessment literacy through the use of the SALRubric.

Table 17: SALRubric Scores: Bella

SALRubric dimensions	Int 1		Int 2	Int 3		Int 4		Int 5
1. Ability to describe assessment	1		2	2		2		3
2. Knowledge of purposes of summative assessment	2		3	3		3		3
3. Knowledge of what to assess	1		3	3		3		3
4. Knowledge of assessment strategies and task design	1		2	2		2		2
5. Knowledge of assessment interpretation	1	Practicum	3	3	Practicum	3	Teaching experience	3
6. NCEA assessment	1		2	2		2		3
7. Preparing students for standards based assessment	0		0	0		2		3
8. Using summative assessment formatively	1		3	3		3		3
9. Understanding assessment consequences	2		2	2		3		3
10. Fairness	1		2	2		2		2

NB. Int 1 refers to Interview 1 etc. (see Table 2 for time frame details)

By the end of the 18-month study, Bella was scored as Competent in eight of the
dimensions of the SALRubric, and at Advanced Beginner level for *Knowledge of
strategies and design*, and *Fairness*. Of interest was Bella's rapid development in
some dimensions early in her ITE year, followed by little perceptible change for a
long period.

The SALRubric scores show that Bella commenced her ITE with limited knowledge about summative assessment and she was judged as Novice for most dimensions, although she scored 2 for *Knowledge of purposes of assessment*, and *Understanding of assessment consequences.* Bella showed significant development in her summative assessment literacy at the second data collection point after her first practicum. There was documented evidence of improvement in 8 of the 10 dimensions, with 3 dimensions being scored at 3 (Competent) surprisingly early when compared with all of the other beginner teachers. However, no further development was noted in these eight dimensions until six months into her first year of teaching.

As can be seen in Table 17 her ability to describe assessment increased after her first practicum, with this development maintained until she had been teaching for 6 months when she moved to Competent status. At the start of the study her description of assessment was:

> It's still something that's marked, and it goes towards getting into NCEA I suppose, but that's about it. [Bella, I1]

However at the end of her university year Bella was able to provide a much more comprehensive description as illustrated in the excerpt below:

> So there's diagnostic, formative, and summative. Assessment doesn't have to be just tests and the internals and the externals, but you can have a lot of informal ones … It can be fun, it doesn't have to be boring, and scary and stressful. You can make them exciting, not really the externals but you can make the internals more exciting as you link them to context so it can be more exciting. Assessments are good, so you can gauge where the students are at and you know where to help them and see where the students need attention. [Bella, I4]

As evident in this quote Bella identified the diagnostic, formative and summative purposes of assessment. She emphasised making assessment in science 'fun'. Six months into her first position teaching she had a more integrated approach to assessment including an understanding of the ethical elements involved.

On the other hand there was no evidence of change in dimensions 7 and 9 (*Preparing students for standards-based assessment* and *Understanding assessment consequences*) until near the end of her ITE programme (I4), after her second practicum. A possible contributor to this could have been Bella's

convictions about teaching to the test, mentioned in three of her interviews (I1, I2, I4). She felt very strongly that anything that could be construed as teaching to the test should be avoided and perhaps this interfered with her development of understanding in these two dimensions.

Bella highlighted *Knowledge of assessment strategies and task design* as an area of weakness. Although she felt she had made some progress, she still lacked confidence in task design at the end of the study. She appeared tentative in her understanding of assessment criteria, and her uncertainty was bolstered by conversations with her ATs in her first practicum school, in which they were very critical of the 'vague' nature of the achievement standard criteria, and of the accuracy of marking and external moderation. In I4 she admitted that she was still not confident enough to design an assignment task:

> Designing them [tasks] I still need a bit of help from colleagues to make it well. [Bella, I4]

Her first school of employment did not require her to design summative assessment tasks, so this may have meant she had fewer opportunities for development in dimensions 2 and 5.

Bella was aware of her development in summative assessment literacy over the course of the study and attributed much of her growth to being able to apply what she was learning at university to a classroom setting on practicum, as mentioned above. She identified specific university classes as having helped her develop her understanding of summative assessment, in particular rubric development in her biology curriculum course, work in literacy, and the opportunity to mark student work in the science curriculum course. However she emphasised that it was through application of what she learnt that she really understood the significance of university course content.

In her final interview (I5), Bella described feeling restricted in her summative assessment literacy development because the school in which she was working did not give her freedom to design and mark her own summative assessments. She observed that in her school formal tests were used almost exclusively for summative assessment,

> It is kind of like it was when I was at school. Here's the topic, here's the test. Not really different [Bella, I5]

Bella had a positive pro-active approach to her own learning and took what opportunities she was given, for example, she did describe her attempts to adapt a Year 9 test on "Living World" in order to make it more inclusive. She felt her confidence as an assessor was still growing, and she looked forward to being given more freedom in the future.

6.2.2.2 Susie

Susie completed a Bachelor of Science (Technology) degree majoring in biotechnology three years before joining the ITE programme. She had worked as a scientist in a hospital setting, diagnosing chromosome disease, in the intervening years. Although she enrolled in ITE and began teaching, Susie's medium term goal was to work in the area of resource development or education policy rather than to work as a classroom teacher. Susie was very confident in her content knowledge and passionate about sharing it with others. She felt it was important that all community members develop scientific literacy.

Susie explained that most of her focus in teaching centred on helping students learn science content in a structured environment. She ascribed to a constructivist view of learning, believing that students need to build on the science they already know. In particular, Susie talked about her students needing to develop an understanding of key science concepts on which to build knowledge:

> You need to understand, or know the process before you step to the next step. Especially in science, you need to know the key concept, you need to understand just some of the key functions in order to understand the next step, because some things you won't get unless you know the basics. [Susie, I1]

She explained the need to scaffold learning for students, and to check for misconceptions of key science concepts. The artefacts she provided included unit plans with well-structured learning sequences crafted so that students could build on their current knowledge and make links with what they already knew through activities and discussion. Susie included a range of visual resources such as video clips, posters, and models in her planning because she believed that science

resources needed to include visual aspects wherever possible. She also believed that repetition helped students gain a fuller understanding of science concepts.

She valued summative assessment tasks for the information they provided to her as a teacher, but also for the learning that could result for students (e.g., as described in Section 5.3.2, Susie designed and used a task on edible vaccines that required students to use science in an integrated way).

Table 18 below summarises Susie's development in dimensions of summative assessment literacy through the use of the SALRubric.

Table 18: SALRubric Scores: Susie

SALRubric dimensions	Int 1		Int 2	Int 3		Int 4		Int 5
1. Ability to describe assessment	1		2	2		3		3
2. Knowledge of purposes of summative assessment	1		3	3		3		3
3. Knowledge of what to assess	2		2	2		3		3
4. Knowledge of assessment strategies and task design	1	Practicum	3	3	Practicum	4	Teaching experience	4
5. Knowledge of assessment interpretation	2		2	2		3		3
6. NCEA assessment	1		2	2		3		4
7. Preparing students for standards based assessment	0		3	3		3		3
8. Using summative assessment formatively	2		3	3		3		3
9. Understanding assessment consequences	1		1	2		2		2
10. Fairness	1		2	2		2		3

NB. Int 1 refers to Interview 1 etc. (see Table 2 for time frame details)

By the end of the 18 month study Susie was scored as Proficient for *Knowledge of assessment strategies and task design,* and *NCEA assessment,* and as Competent in all other dimensions except for *Understanding assessment consequences*, where she was scored as *Advanced Beginner.* Of interest was Susie's rapid development in task design including NCEA tasks (usually seen as technically challenging).

Susie was scored as a Novice for most dimensions of summative assessment literacy at the commencement of her ITE, scoring 2 for *Knowledge of what to assess*, *Knowledge of assessment interpretation,* and *Using summative assessment formatively*. She showed an increase in her competency for most dimensions by the second data collection point which was after her first practicum, and an increase in the dimension *Knowledge of assessment interpretation* at the third data collection point. She had not made progress in this dimension between interview one and two. At the fourth data collection point, Susie showed shift in five dimensions and was scored as Competent or Proficient for all but two dimensions. At this stage she talked

confidently about being able to summatively assess students in her first teaching position. In this, Susie's profile is substantially different from that of Bella whose comments suggest she had a much simpler understanding in areas of assessment strategies and task design, and NCEA and standards based assessment at Interview 4. Unlike Bella, Susie made noticeable gains in a number of dimensions after each practicum experience, and was particularly capable in the areas of NCEA assessment design.

By the end of the study, Susie was the highest scoring of the beginner teachers overall, with the only dimension below Competent being *Understanding assessment consequences*, which was commonly low for all beginner teachers.

Susie entered ITE with a very positive approach to summative assessment and a strong belief in herself. Although she admitted she was quite outcomes-focussed and a high achiever herself, she very quickly became aware of the diverse range of students that she was teaching, and that not all students approached assessment the way she did. She realised she could give her students options within assessment tasks, so that they could choose or negotiate a task that suited their strengths:

> I'm glad it's not just test stuff, and that you've got a bit of leeway as to how you want to design it. For example in the senior biotech one, basically there are these two options, which one do you want to do? And we came up with a report supported with a poster, well it ended up being just supported with diagrams because it ended up being easier that way. But actually having a bit more freedom. [Susie, I4]

Her comments indicated she aimed to respond to the needs of her students. As another example, she explained a unit plan that she produced as an artefact in Interview 4, and commented on how she had adapted the school-planned summative assessment task to include oral presentations when she realised that some of her students had difficulty with writing reports. Students were given the option of a written or oral report.

> So is just getting to know your students, and designing your assessments around the students. You still need to follow the guidelines of the rest of the class, but adapting it slightly. [Susie, I4]

It was clear from this comment that she had the confidence to adapt school-designed summative assessment tasks to better suit her students. Similarly in Interview 5 she talked about using this flexible approach to task design in her school:

It's what your students are interested in, and design them to that. Like with the posters, with the year nines, they really enjoy that and I'll actually sit down and do something like that. They are quite a creative sort of bunch see you design [assessments] more to be like that. They are still applying their knowledge. [Susie, I5]

By the end of the study Susie was talking confidently about using a responsive approach to summative assessment, and designing and using science summative assessment tasks at all levels of the school:

I am comfortable enough to be able to design an assessment task like an NCEA achievement standard, write that out and get it approved, and follow all the way through without any issues. And then I've got internal marking and moderation … And you don't seem to be any issues at this stage. My judgements seem to be on par with others, if I have questions I ask. I think I'm getting there. [Susie, I5]

Susie was fortunate in being placed on practicum in two schools that permitted her considerable freedom. In her first practicum she was able to design and mark a Year 9 summative assessment activity as well as play a role in a number of other assessment events. Similarly in her second practicum Susie was given multiple opportunities to be involved in assessment, for example she marked examinations, took an active role in moderation meetings, supervised re-sits for NCEA, and designed new tasks. The school in which she gained her first employment as a teacher also gave Susie relative autonomy when it came to implementing summative assessment for her classes, which she said allowed her to continue her development of summative assessment literacy.

Bella and Susie's school experience differed considerably during ITE and in their first schools. Bella was more restricted in each context, and therefore had fewer opportunities to develop her confidence and practical knowledge. During her ITE Susie was placed in schools that encouraged her to experiment with assessment types, and she relished these opportunities, and produced artefacts of tasks including unit plans, lesson plans, worksheets with links to websites, tests, and posters, most of which she had developed herself. These showed evidence of close attention being paid to detail and were of good quality.

6.2.2.3   Ryan

Ryan entered the ITE programme after many years working in construction in Australia. His success and satisfaction in training apprentices motivated him to consider a career in teaching. On returning to New Zealand he completed a Bachelor of Science degree with the plan of becoming a secondary science and technology teacher.

Ryan prioritised the relationships he had with students and invested time in developing positive relationships with them. He had a holistic approach to learning, looking at the student first as a person and then considering the science content he had to teach.

> I'm not saying exams per se are bad, but they need to be incorporated more into a holistic view of the subject. [Ryan, I2]

He believed that when students were able to relax and be friendly with their teacher they would be more likely to be motivated and focussed in class. He explained that he worked on using a relaxed but purposeful teaching style. Ryan believed that another important element to motivating students was to provide links between their schoolwork to the outside world. In this way he could be seen to have a constructivist approach to teaching and learning. He explained that it was important to build on students' knowledge and strengths, relating new knowledge to their current understandings of the world. He felt that it was very important for students to understand the 'big picture' of their learning. Ryan spoke of how he kept students motivated by telling stories, and linking what they were learning with practical applications of that knowledge:

> … and then if I come up with a story from something that I did in the industry all of a sudden they are rapt and they are full of attention. If they are interested, then that is the key. [Ryan, I5]

He also believed that practical activities and the use of computer technology were ways to engage learners. He voiced strong opposition to the use of external examinations, as he felt they did not help students learn and did not promote deep learning. He believed that examinations encouraged rote learning, which he felt reduced students' engagement and hindered real learning. Ryan entered ITE with a

very negative view of summative assessment based on his personal experiences and emotional responses to being assessed.

> Again speaking from my own experience, and going back further, it was all learning by rote. I hated it, and I got good at doing it... So you got your marks, and you passed, and whatever you had to do, and then you just forgot about it... from my own personal experience, mainly exams, and as I say I hated them. I got stressed out, wouldn't sleep, all that stuff. [Ryan, I1]

Ryan explained that these experiences meant he wanted to teach in a way that did not emphasise formal assessment. Ryan had gained experience of assessing others in a practically based career before entering ITE, and he felt that the methods used for assessment in that context gave him insights and strength in alternative forms of summative assessment.

Table 19 below summarises Ryan's development in dimensions of summative assessment literacy through the use of the SALRubric.

Table 19: SALRubric Scores: Ryan

SALRubric dimensions	Int 1		Int 2	Int 3		Int 4		Int 5
1. Ability to describe assessment	1		2	2		2		2
2. Knowledge of purposes of summative assessment	1		1	2		2		3
3. Knowledge of what to assess	1		1	1		2		2
4. Knowledge of assessment strategies and task design	1	Practicum	1	1	Practicum	2	Teaching experience	2
5. Knowledge of assessment interpretation	2		2	2		2		2
6. NCEA assessment	1		1	1		2		2
7. Preparing students for standards based assessment	1		1	2		2		3
8. Using summative assessment formatively	2		2	2		2		2
9. Understanding assessment consequences	2		2	2		2		3
10. Fairness	1		1	1		2		2

NB. Int 1 refers to Interview 1 etc. (see Table 2 for time frame details)

By the end of the 18 month study Ryan was scored at Advanced beginner for seven dimensions and Competent for three dimensions. Of interest was Ryan's slower rate of shift, with only one dimension showing shift after his first practicum, in contrast to the rest of the beginner teachers. More shift was evident in Interviews 4 and 5 than earlier in the study.

At the beginning of the study Ryan was scored as Novice for seven of the 10 dimensions and Advanced Beginner for the other three. Ryan showed very slow

development in almost all dimensions, and by the end of the ITE year, he scored Advanced Beginner for all dimensions. At Interview 2 after his first practicum, Ryan showed shift in one dimension *Ability to describe assessment.* After his second practicum at Interview 4 he showed shift in four of the 10 dimensions. At the final data collection point Ryan showed further shift in three dimensions, and scored 3 (Competent) for 3 of the 10 dimensions. For two of the dimensions, *Knowledge of assessment interpretation* and *Understanding assessment consequences*, Ryan showed no measureable development over the entire study.

Ryan's views included scepticism of examinations and a lack of faith in internal NCEA assessment because he perceived teachers were teaching to the test. During his ITE, Ryan was placed in two practicum schools with a reputation in their districts for their conservative approaches to summative assessment. He reported he was given very little control over the ways he was able to use summative assessment and this reinforced his negative views. In both practicum schools Ryan reported what he saw as assessment-driven teaching:

> I'll … say I haven't changed, if anything I'm probably more determined in my views about assessment. Teaching and catering just for the exam was blatant and it was ridiculous. They might as well have told them what was in the exam. [Ryan, I2]

> It was very similar to the first practicum … very much structured around the exams and teaching to the exams basically … it is too heavily emphasised and it encourages rote learning that I know doesn't work. [Ryan, I4]

He felt his university courses gave him a basis from which to view summative assessment but his practicum experience did little to help him prepare for assessing students in the way he felt was best.

However, in his first teaching position Ryan had more freedom to choose how to assess students, and the resultant confidence and development was evident in Interview 5. Although he still perceived that there was a big emphasis on the gathering of NCEA credits, he did appreciate what he was able to learn from his mentor teacher:

> But he [mentor teacher] is quite proactive and doing his own assessments, which is a good learning experience for me … I've pretty

much got an open book and I can do what I want to in the classroom so I have designed some [assessment tasks]. [Ryan, I5]

Ryan produced a series of worksheets and a final assessment task to illustrate what he had started doing in his school.

Because of Ryan's assessment beliefs it was unfortunate that both of his practicum schools reinforced his beliefs about teaching to the test because of their focus on formal summative assessments leading to NCEA credits. These school provided him with very little opportunity to be involved in summative assessment.

These three vignettes give an indication of the variety of developmental patterns for individual teachers. Idiosyncratic patterns were evident in teachers' development of assessment literacy over time, as shown in Figure 4. Bella was the least confident of the beginner teachers, but was keen to learn from practicum. Her first practicum appeared to act as a catalyst for her summative assessment literacy development, and after that her development was gradual. Ryan also lacked confidence and also had a very negative view of formal assessment. Like Bella, both of his practicum schools gave him little scope to develop and use summative assessment. His development as measured by the SALRubric was the slowest of the beginner teachers. Susie was the most confident beginner teacher, and each of the schools she worked in gave her considerable experience with summative assessment. By the end of the study she was the most confident about summative assessing students. Although these three teachers experienced mostly the same ITE programme, their practicum schools were quite different, and apparently as a result their developmental trajectories are quite different.

The following section summarises the overall patterns evident within SALRubric dimensions from all beginner teachers in this study.

6.2.3   Summary of patterns of individual development

The scores generated from the SALRubric (see Appendix H for full table of results and Figure 4 for their graphical representations) draw attention to a number of features of summative assessment literacy development for the beginner teachers in this study. All teachers shifted their scores over time but to varying extents. Although each beginner teacher in this study presented his or her own

unique development pattern, common patterns of development were evident across teachers as well.

- Six teachers in this study (Bella, Elisa, Kate, Mary, Susie, and Wiremu) demonstrated a shift in all 10 dimensions of assessment literacy from lower to higher expertise; however, Ryan showed a shift in 8 out of 10 dimensions, and Benjamin in 9 out of 10 dimensions.

- For all teachers, with the exception of Ryan, the main shifts were noticed between Interview 1 and Interview 2, with further shifts occurring between Interview 3 and Interview 4. Between Interviews 1 and 2, and then Interviews 3 and 4, all teachers experienced practicum. Little shift was noted between Interview 2 and Interview 3, during which time the teachers were completing university courses.

- All teachers, with the exceptions of Susie and Wiremu, made shifts in 3–5 dimensions between the last two data collection points (Interviews 4 and 5), i.e., between the end of their ITE programme and after six months teaching in their first schools.

- For each individual a distinct pattern of development was observed. Development occurred across a number of dimensions at the same time but at different rates

- Some teachers (for example, Elisa and Kate) demonstrated shift more quickly than others (for example, Benjamin and Ryan).

- For all teachers, except Benjamin and Ryan, a shift of two stages for at least one dimension was achieved over one time period.

- For all teachers except Mary, initial appraisals indicated they were above Novice for some dimensions of assessment literacy.

- No teacher reached Expert level for any dimension within the 18 month time frame of this study, and none remained at Novice.

In summary, the patterns of shift in dimensions of summative assessment literacy were idiosyncratic for beginner teachers but the greatest shifts were noted after the beginner teachers spent time in schools (either on practicum or in their first school of employment). Some teachers made shifts more quickly than others, with Ryan making the least gains over the period of this study, and Susie showing the

greatest gains. Shift in some dimensions seemed to be slower than others for all teachers.

6.3   Contributors to development over time

The findings in this study clearly show development in dimensions of summative assessment literacy over time for all beginner teachers. The beginner teachers were asked if they could identify particular contributors or drivers that challenged or shaped their thinking, promoting their development over the duration of the study. All teachers in this study identified specific people (Section 6.3.1), learning activities (Section 6.3.2), and teaching experiences (Section 6.3.3) they felt particularly helped in their development of the understanding and practical use of summative assessment.

6.3.1   People

Beginner teachers in this study spoke of the contributions of people to aspects of their development of summative assessment literacy, particularly specific university lecturers, associate teachers and classmates.

Firstly, all beginner teachers in this study identified lecturers from their ITE programme who helped them learn about summative assessment. They reported that the lecturers in curriculum papers (Lecturers X and Y) were particularly helpful in this regard throughout their year of ITE. One literacy lecturer (Lecturer Z) was also highlighted by all beginner teachers in this study as having considerable input. These identified lecturers helped the beginner teachers make links between the New Zealand curriculum, their planning, teaching and assessment, and helped them to grasp the meaning of summative assessment within the 'big picture' of learners' progress. The teachers talked about the effects of getting to know these lecturers and using them as models for their own teaching. There was an obvious personal connection made between the beginner teachers and the lecturers, and these lecturers seemed to have a much greater effect on the teachers than others:

> I think actually the biggest help to me has been the curriculum papers
> I've been doing, and the teachers of those papers. I have had two
> really excellent teachers, and I've learnt more from them. Frankly I
> don't think I've learnt much at all from the core paper ... So I would

say basically Lecturer X and Lecturer Y have been a huge help with my learning. [Ryan, I3]

When I was in Lecturer X's class I learnt a lot about assessment … I think with Lecturer Y it was really good, plenty of ways to assess, she was really good. [Benjamin, I3]

A lecturer who taught a module on literacy across the curriculum was singled out by six of the beginner teachers as being extremely influential. Although the main focus of his workshop series was not summative assessment, the lecturer wove in aspects of assessment for teachers' consideration. He introduced a wide range of literacy tools in his course and introduced the concept of ecologically valid assessment, which was memorable for them. The beginner teachers in this study related to him well and appreciated his relaxed manner, and they commented positively on this input. This social relationship seemed integral to beginner teachers learning about summative assessment and was long-lasting.

Yes, but in Lecturer Z's class he always told us about how you need to assess like you teach. And I don't think you see that very often in schools. [Mary, I2]

Yes, for Lecturer Z's one I think … that was just … I'm still thinking about how to assess thinking and tools, so that I think was a big thing. [Benjamin, I3]

Lecturer Z was still specifically referred to by the beginner teachers in the year following their ITE. It appeared that beginner teachers were looking for a personal connection with lecturers and when this was developed they were more open to learning from them. Of all of the lecturers the teachers were exposed to, Lecturers X, Y and Z were mentioned frequently by everyone. The teachers spoke of these lecturers with fondness and gratitude, emphasising their knowledge and also their way of relating to the ITE students.

Practical teaching experience involved beginner teachers in this study with a large number of other social interactions, as they worked with their ATs, colleagues and mentors as well as students. The beginner teachers had varied experiences when on practicum. A number spoke of the contributions of their ATs to their development, both positive and negative. However, all the beginner teachers regarded their practicum experiences as having helped them develop as teachers,

and most returned to university after practicum experiences feeling buoyed up by the successes they gained through opportunities they were given. Some ATs were very generous with their time when working with the beginner teachers but others were not. The contribution of ATs was described on a continuum from very positive to very negative. The beginner teachers spoke at length about the interactions they had with their ATs and how much they learnt from either observing their teaching or getting feedback on their own work. All beginner teachers in this study had engaged with this feedback and were able to describe ways they either would, or would not, run their classrooms based on the practice of their ATs:

> Yes, where we get ideas from the teachers (ATs) because they actually do it. [Bella, I3]

> I would say just seeing, and just carrying it out in classrooms, and seeing how teachers deal with it, the process they go through from teaching to be training the students for what they are going to be assessed on, and then just seeing that whole different range. And then knowing that the experienced teachers are sometimes disappointed in the results. [Mary, I4]

What they were told by their ATs was trusted more highly than what they heard from university lecturers because the ATs were perceived as having more relevant and up to date knowledge than the lecturers.

Beginner teachers in this study found that their classmates in the ITE programme were another source of help and support. For example, when working through the process of peer-critiquing summative assessment tasks they found that working with their classmates was invaluable in helping them understand the content:

> Yes, it was good when Lecturer X would give us lots of standards, and would talk through them ... It's really good how she gets us to work in partners, because you still have ideas yourself but if you're paired up with someone you can make up a better idea with them, so that's quite good. Everyone in that class has really good ideas about different things, I like working with all of them. [Bella, I3]

Their social interaction with classmates helped the beginner teachers develop a more in-depth understanding of summative assessment. In a large number of the interviews, beginner teachers in this study commented on the value of working in

small groups or in pairs, as they felt this collaboration contributed to their learning.

6.3.2 Learning activities

Beginner teachers spoke candidly about how the process of being involved in specific learning activities within the ITE programme contributed to their development of summative assessment literacy (e.g., structured curriculum workshops, designing unit plans, rubric module, literacy across the curriculum workshops, marking student work). They also provided commentary on the links they made between the university courses and the practice of classroom teaching, and they discussed the benefits and limitations of the university courses they were involved with. A small number of activities were identified as being particularly helpful in teachers' summative assessment literacy development. These activities helped the beginner teachers make links between the curriculum, planning, teaching and assessment, and this 'big picture' enabled teachers to grasp the meaning and uses of assessment.

Firstly, a number of beginner teachers talked about the helpfulness of the structured curriculum workshops, which included lesson planning and unit planning. The beginner teachers were forced to consider how they would know whether students had learnt the science being taught. The workshops included backward design, working backwards from identified achievement objectives and end-points to finding appropriate summative assessment tools, learning materials, teaching sequences and so on. This process of identifying learning objectives and then planning to measure whether students had met the objectives helped teachers understand the rationale for assessment and its usefulness:

> Yes, and we've just started doing a unit plan with Lecturer X getting an internal [achievement standard] and looking at it and thinking, what exactly do they need to know from that. And then how are we going to put that into our unit plan, not just a question/answer, this is the information you need to know, how do we intertwine it to an assessment? So I guess going backwards from what they actually need to know and creating lessons, I feel more confident. [Mary, I2]

> Because when we were designing the unit plans it was quite helpful, as you design the sequence yourself and then you follow it through and make sure you have covered everything before you give them the

213

assessment. So I found that was quite helpful just to link everything right through to the end. [Bella, I3]

By working on a unit plan for teaching activities in university classes, the teachers carefully considered how to design summative assessment tasks that would measure whether students had met the learning outcomes. In doing so, the teachers learnt about issues of validity and reliability and were able to more clearly define learning outcomes. The structured workshops also gave the teachers more confidence with their planning overall as they felt they were planning from a better foundation.

Secondly, a specific short module on rubrics taught within a university course was identified as being particularly useful in the development of summative assessment literacy of teachers in this study. Not all beginner teachers were enrolled in the course where this module was taught, but those enrolled emphasised the effectiveness of the tasks. The rubric task involved teachers taking the achievement criteria from an achievement standard task and developing a marking rubric for an NCEA assessment task, thus providing more specific criteria to allow for judgements to be made. The teachers then used the rubric they had developed to mark real student work. They found the learning activity particularly powerful.

> Well, after she had given us the standards she gave us the assessment schedule, and she gave us about four different samples of work that students had done. We had to make rubrics from the assessment schedule just to make it easier for us to mark … So we just made them and then marked the different pieces of work, and I thought that was really helpful. [Bella, I3]

> I think the tasks we did with Lecturer Y about creating a rubric from the criteria itself … Because when you create the rubric you kind of set it up so it is almost a checklist, and your rubric will take into account variable answers, whereas the actual criteria [from the achievement standard] doesn't, and there is a lot of stuff that is open for interpretation. Like I said, the template that Lecturer Y gave us for creating a rubric from a criteria … that for me has been the single most helpful thing because it has given me focus. [Ryan, I3]

The rubric module helped the beginner teachers understand the nuances of criteria-based assessment and the care needed when making judgements about student work.

214

A third highly effective learning activity that was identified by beginner teachers as contributing to their developing summative assessment literacy was a series of workshops focusing on literacy across the curriculum. The teachers found this short series particularly useful, partly because it included a large number of examples of strategies they could use in the classroom to assess student learning, and partly because the lecturer linked ideas about literacy, learning and assessment in ways the teachers could understand and appreciate. Most teachers had identified literacy as a challenge for some secondary students and so they were aware of potential fairness issues that could come up in science summative assessment. In particular the input from the series on ecological validity was commented on and did influence the teachers' practice. Mary commented on the need for ecological validity, but its scarcity in secondary schools:

> In Lecturer Z's class he always told us about how you need to assess like you teach. And I don't think you see that very often in schools. [Mary, I2]

A fourth highly effective learning activity was practise in collaborative marking of real students' work using prepared materials such as marking schedules. This was extremely useful in both helping teachers develop a good understanding of the requirements of particular achievement standards and forcing the teachers to make judgements about students' work. Tasks which focused on marking work that related to NCEA standards were mentioned in particular. The marking was completed collaboratively with classmates and this interaction helped everyone in the group learn how to apply the assessment criteria and be able to justify their grades much more clearly and accurately:

> Well, just the realisation of how important moderation is [when marking] because when we judged them then we compared them to our classmates. And I think that it may also be the reason why I thought the technique and the task was so helpful. Because afterwards we debriefed with each other and we collaborated in that made things … Even when you were unsure about something … When others would speak about it, or were unsure about it, it made you realise something you could look into more … Or if others could quickly answer something and put at ease all of your doubts, then it was okay. [Wiremu, I3]

> It's really good how she gets us to work in partners because you still have ideas yourself but if you're paired up with someone you can

make up a better idea with them, so that's quite good. Everyone in that
class has really good ideas about different things. I like working with
all of them. [Bella, I3]

The collaborative marking was found to be extremely helpful for those teachers
who were able to do NCEA marking on practicum, as the experience gave them
confidence.

6.3.3 Teaching experiences

All beginner teachers felt that their practicum experiences were crucial to their
development as teachers and helped them develop their summative assessment
literacy. They all returned to university after practicum experiences feeling more
confident in aspects of their summative assessment practice. For example, in this
quote Elisa explained how the process of teaching a unit and then testing the
content helped improve her confidence:

I felt better because when you teach the unit you know what you are
marking for, and what you're looking for. And even when you make
the test you know exactly what you're looking for, you know how to
mark better. You don't always have to refer to another bit of paper
with all the answers on it because you know it straight away. It's just
the practicum. [Elisa, I4]

The beginner teachers who were able to use formal summative assessment tasks
while on practicum were appreciative of opportunities they were given. However,
not all beginner teachers were given the opportunity to develop and use
summative assessment tasks on practicum even after teaching whole topics in
some cases. These teachers talked of not being allowed to mark work or design
tasks. They felt the opportunity to formally assess students would have been very
beneficial for their summative assessment practice. They felt that early
opportunities would have been very beneficial to their summative assessment
literacy development and felt aggrieved that their ATs did not permit them to
assess the students, as illustrated in these quotes:

Because I didn't really get to do any summative assessment, other
than the feeling I get when I came away [at] the end of the lesson
when I was thinking 'has it sunk in?' [Ryan, I2]

In my last practicum I really didn't do any kind of marking or
assessments … Because I asked to do some marking for the one I was
there for, but she just said "Well, the other teacher's doing it so you

216

can go along with him and help mark." But the other teacher I wasn't really working with, and it would just be annoying for me to show up and he wasn't ready for it. [Mary, I3]

Five of the beginner teachers discussed how practicum and teaching allowed them to consolidate their learning from university courses. Bella explained that it is by doing things she really gets to know them:

I think, even if I have already learnt it, I probably won't remember it until I'm in the classroom. I kind of have to practise something to know it. [Bella, I1]

Others talked about the impact of the combination of theoretical learning from university with practicum. They saw these learning opportunities as being complementary:

We had a couple of lectures [about summative assessment] this year before I went [on practicum], but I sort of learnt about it on-the-job, and then saw the importance myself. [Bella, I2]

I think it is just a combination of practicum and uni, seeing the different ways you can do it and that it can be done efficiently. [Susie, I4]

Wiremu talked about the way the assessment theory he learned at university raised his awareness and meant that he noticed things he may otherwise have overlooked when on practicum. He felt that by having his awareness raised in university courses, he was able to gain more from the practicum experience, as illustrated in this quote:

Well, I guess when we were doing the lectures that's when I started getting quite introspective about what I thought about assessment … It made me aware that I had to think about these things, and then so when I was on prac and something assessment related would come up I wouldn't just overlook it. I would be like – "stop and pay attention to this because you're not satisfied with your perspectives on assessments yet." It's making me more aware of how important it is, so things when you might normally overlook something, you recognise it and kind of just put it in the back of your mind. I'll think about this later on. [Wiremu, I2]

Similarly, Susie found that a heightened awareness of issues to do with assessment through exposure to ideas at university helped her once she was on practicum.

However, when the beginner teachers began working in their schools they recognised that they were continuing to learn about summative assessment, although four of them were not given the responsibility and freedom they wanted in their first year of teaching. The teachers who were able to take responsibility for summative assessment talked about their increased feelings of confidence given that they now had responsibility for assessing their own students' work:

> [I'm feeling] much more solid than I did last year. Actually doing it and having to do it, having to mark it and moderate. It's good. [Bella, I5]

> I feel much more assessment capable than I have, every week previous to today. I feel that progress is weekly. It's scary how much I feel I've grown … Because that's when you get all the hands-on experience… Things I learned last year were very very different to what I learnt in my science degree. It wasn't so much about knowledge, it was about people. And personally I found that amazing. It was awesome because I dislike treating people like numbers. [Wiremu, I5]

There was a general consensus from beginner teachers in this study that the university classes in their ITE programme were useful, but that the time spent on practicum in schools was more useful. A number of teachers had advice for the university programme. They felt that general principles of assessment should be a teaching focus in the generic compulsory courses, which at the time of the study only provided very basic input about assessment. Teachers felt this would have given them a sharper focus on assessment, especially if a compulsory assignment on this content was included.

In considering his learning about summative assessment, Wiremu made some interesting observations. He talked about the contrast of being assessed by the university as a preservice teacher and his role in assessing others whilst on practicum. Here he made the important link between one of the effects of assessment: the fact that what is assessed is seen as more valuable or important by those who are being assessed. Because his abilities as an assessor were not assessed by the university during his teacher education, Wiremu felt that beginner teachers did not prioritise their own development of assessment literacy.

In summary, this section highlights the contributors to beginner teachers' development as identified by the beginner teachers. In some instances the same

specific person or specific event influenced every teacher in this study. Elements from both the university programme and practical teaching were acknowledged as having an effect on the teachers' development of summative assessment literacy. All beginner teachers identified specific lecturers from university courses and teachers from schools who helped them to develop a wider understanding of summative assessment through the learning activities they provided and through their own supportive relationships. Beginner teachers in this study identified four university-developed learning activities that had a profound effect on their development of summative assessment literacy. These activities involved unit planning, rubric development, literacy across the curriculum classes and marking students' work. The teachers' experiences in classrooms were also identified as having a major effect on their development.

6.4   Summary

This chapter has presented findings about the development of beginner teachers' summative assessment literacy over the timeframe of the study. Many of the dimensions in the SALRubric clearly linked with earlier literature describing assessment literacy in science teachers (Abell & Siegel, 2011; Lyon, 2013a) and others related to the context in New Zealand. The findings provide evidence of the teachers' development in 10 dimensions of summative assessment literacy. The beginner teachers showed development in all dimensions, with most shift being evident after practicum or teaching periods. There was some evidence of patterns of development across teachers in this study; however, for each teacher the development was idiosyncratic. By the end of their first six months of employment none had reached Expert, and none remained at Novice in dimensions of the SALRubric. The teachers were able to identify particular people, learning activities and teaching experiences that contributed to their development.

The next chapter, Chapter 7, describes the findings relating to four amplifiers and filters, which individually and collectively influenced teachers' summative assessment decision making and practice.

Chapter 7: Summative assessment literacy: Amplifiers and filters influencing assessment decision making and practice

Having knowledge and understanding of the key aspects and principles of sound summative assessment are important characteristics of beginner teacher summative assessment literacy. However, knowledge alone is not enough. Teachers need to embrace, reject, modify or adapt their knowledge to the particular situation they find themselves in and use this situated personalisation of knowledge to inform their assessment decision making for action. The personalisation of knowledge depends on the various lenses, termed amplifiers and filters individual teachers employ (Gess-Newsome, 2015). Amplifiers and filters influence the way beginner teachers use knowledge about summative assessment in the context of their teaching.

In this study four amplifiers and filters were identified, and these are described in this chapter. These are sociocultural context (Section 7.1), teachers' own conceptions of assessment (Section 7.2), ethical matters related to assessment purposes, enactment and consequences (Section 7.3), and teacher emotional responses to assessment in action (Section 7.4). It was clear in the data that the amplifiers and filters overlapped and affected the teachers concurrently but in this chapter each is foregrounded and discussed in turn, in order to obtain a better understanding of how each influences beginner teachers. It is acknowledged, however, that in practice they may all be involved and in play at once.

7.1 Sociocultural context

The first group of amplifiers and filters relate to the cultural and social contexts of assessment. Assessment as a cultural and social practice has, at its heart, a recognition that assessment occurs within a wider social context, and is not independent of or unaffected by this context. The social origins and consequences of assessment are wide ranging and influential. The beginner teachers in this study identified two elements of the cultural and social nature of assessment that affected their approach and decision making with respect to summative

assessment: school assessment culture (Section 7.1.1), and students' cultural and ethnic backgrounds (Section 7.1.2).

7.1.1   School assessment culture

School culture is reflected in school priorities and ways of doing things, and arises from expectations of school communities (parents, families and other stakeholders) as well as embedded practices. All eight beginner teachers described how the culture of the schools in which they worked influenced their approach to assessment and their assessment decision making.

Beginner teachers were able to identify aspects of school culture that affected their assessment practice. Their comments indicated they had experienced two types of schools when on practicum and during their first six months as teachers: schools which tightly controlled or prescribed what beginner teachers were able to do for summative assessment, and schools which gave beginner teachers considerable freedom. Of the beginner teachers in this study, four commenced teaching in schools where they reported that they had little or no say in how summative assessment was designed or conducted. For example, Benjamin described his school's culture for assessment as allowing him no opportunity to have input into the timing of summative assessments nor the tasks that were used:

> Well, we have a topic that lasts about six weeks and we have a mid-topic test and an end of topic test… it is just this department-wise. Everyone does the same thing … Sometimes I'd like to bring the assessment forward, but then it depends on what the department's schedule is … With my Year 11s we were ready about two weeks before the assessment date was. It was just two weeks of doing the same old thing, and I had to create new stuff every lesson to fill in time [as we waited for the assessment date]. [Benjamin, I5]

In tightly controlled schools, beginner teachers considered they were sometimes forced to act in ways that were contrary to their own beliefs. For example, Ryan was a teacher who believed that examinations could cause students to become disengaged and so he considered it was important to minimise the use of examinations and allow for other forms of summative assessment. However, both of his practicum schools and the school where he commenced working all highly valued examinations, and this required him to use examinations with his classes. He did not enjoy this aspect of the school culture:

> Oh, well, the school I was at, they made no secret of what they were.
> One teacher said to me that the school was assessment oriented,
> meaning exams and NCEA levels oriented. They made no bones about
> that at all. [Ryan, I4]

Ryan explained how this caused him considerable angst as this school culture acted as a filter, narrowing his options as it required him to make particular assessment decisions. His school made no secret of being a school which was examination-orientated, and he had to comply with this philosophy while working there.

As a contrasting example, Elisa worked in a school where she was allowed to make decisions regarding summative assessment albeit within a supportive framework. She had freedom to teach and assess in ways she felt most appropriate, but met with other teachers in the science department to map out their assessment in a flowchart.

> I've got complete freedom basically with these classes, so with the
> Year 11s we had four topics within our unit so at the end of each of
> those we'd do a bit of a flow chart together, and I'd [prepare a task]
> that they'd have to do. Instead of doing it all and assessing at the end,
> doing bits in between, which is really good. [Elisa, I5]

Elisa found that in this school, assessment was treated very casually. On one occasion she decided to redevelop a test for a lower ability Year 10 class only to be told, "You don't have to even test if you don't want to, it doesn't really matter" [Elisa, I4]. The relaxed attitude taken towards assessing students who were struggling influenced Elisa's own assessment decision making processes, as the school expectations filtered or reduced the requirement to assess all students. She was frustrated and worried by the fact that she was told that it did not really matter, leading to what she deemed a casual approach to summative assessment, and as a result she felt less confident in following her own beliefs

Beginner teachers identified that the New Zealand qualifications system contributed strongly to the educational environment and school culture. Because of the nature of the ASs and USs used for NCEA, a common observation by the beginner teachers was that schools delivered the curriculum as content 'packages' relevant for the set of chosen standards for each subject, and they were expected to follow this pattern. The practise internal assessment tasks took up considerable

classroom teaching time, and in some cases the beginner teachers thought they were asked to teach too closely to the real internal achievement standard tasks or as they put it "teaching to the test". They were clear this was contrary to their beliefs about what constitutes quality assessment practice, but that they were likely to teach this way because it was a school expectation.

Another NCEA-linked feature evident in school culture according to the beginner teachers was the "washback effect" of NCEA. This has been observed by the teachers in the summative assessment tasks set for Year 9 and 10 classes. All eight teachers in the study found that at least one of their schools used the NCEA grading system of Not Achieved /Achieved /Merit /Excellence grading system for junior classes. They noted, however, that the tasks to which this grading system was applied were not always linked to any sort of standard. In this situation the beginner teachers had to conform to the school's norms for grading.

Wiremu's experience provides a compelling example of the amplifying influence of school culture on teachers' decision making. Early in the study, Wiremu voiced his strong belief in using a holistic approach to assessment. However, after spending seven weeks on practicum at a school well-known for using a particularly structured and rigid system of assessment, Wiremu's views appeared to change significantly. He became far less flexible in what he identified as good assessment practice, and shifted to a philosophy which espoused the use of a narrow range of assessment types. As mentioned in Section 5.3.2, he commented in the second interview as follows:

> That's not saying we shouldn't think of ways to target children that would have a fairer assessment by not doing your standard written tests, but I guess the bottom line is that when it comes to it, I don't want to say 'being a scholar' but that's the only thing that is coming to mind. If you want to be a scholar you have to be tested like scholars are tested ... You can't just be outside of the system, and be included still. [Wiremu, I2]

In this case the school culture was evident as a strong change agent on Wiremu's summative assessment decision making. His time in this particular school changed the basis on which he made assessment decisions. However, in his second practicum school (with a more relaxed approach to assessment) Wiremu

changed back and focused on holistic assessment again. In Wiremu's case the culture of the school greatly affected his approach to assessment.

As discussed in Section 6.3.1, beginner teachers' ATs and mentor teachers at school influenced their practice and confidence levels. The interactions they had with these people encouraged them to view assessment in particular ways. Their involvement in social moderation for NCEA would be seen as especially important, although by the time when the study ended many of the beginner teachers were only starting to be involved in that process.

Overall, in several instances school culture or norms of operation acted as a strong filter and amplifier for beginner teacher summative assessment decision making. In some school situations they could not make decisions to act in ways they wanted to, and in others they could. In Wiremu's case the school culture was seen to change the basis on which he made assessment decisions.

7.1.2 Students' cultural and ethnic backgrounds

Consideration of the cultural backgrounds of students was a dimension influencing teacher summative assessment decision making. All teachers wanted to teach and assess in ways that were responsive to students from a wide range of cultures. They talked about the need to consider the range of cultural and ethnic backgrounds represented by students in their classes. The teachers themselves represented a number of cultures, in particular New Zealand Pākehā, New Zealand Māori, and European. They were aware of differences between cultures and were particularly aware of the needs of Māori students, and students with Pacific Islands heritage; two groups of students identified as priority learners in New Zealand.

The most striking example of this focus came from Wiremu, a Māori teacher. As noted in Section 5.5.2, Wiremu felt that the mode of assessment (whether it be written, oral, or visual) needed to be considered when planning an assessment. He valued oral assessment in Māori-medium schools. However, he was aware of his own lack of cultural knowledge when it came to other groups of students such as Somalis, Chinese, Indians and so on, particularly when he was on practicum in a large urban multi-cultural school (I2). Wiremu felt his lack of cultural knowledge

restricted his ability to assess in the best way possible for particular groups of students.

Kate and Bella also expressed a lack of confidence when it came to making decisions about assessing students from cultural backgrounds different from their own. They did understand the issues facing students from a range of cultural and linguistic backgrounds, but did not necessarily know how to meet their needs. They wanted to learn more about how to assess in culturally responsive ways:

> But not [I don't have confidence] in… Should we be assessing them differently, should we be assessing them the same, should we be changing the assessment, should we not be changing the assessment? [Kate, I3]

Knowledge of students' cultural and ethnic backgrounds was a lens these beginner teachers' were aware of when they were considering how to assess their students, although at this stage in their careers they were very uncertain about how best to do this.

In summary, findings have been presented regarding the amplifying and filtering effects of social and cultural environment on teachers' summative assessment decision making. School culture, especially the specificity of their approaches to summative/ NCEA assessment, and the ethnic and cultural backgrounds of students, were said to influence teachers as they considered how, when and why to assess students. The following section presents findings on the second group of amplifiers and filters: teachers' conceptions and background experiences.

7.2   Conceptions of summative assessment

Beginner teachers in this study expressed ideas and provided artefacts that signalled their conceptions of learning and assessment. These conceptions appeared to be deep seated and acted as amplifiers and filters for teachers as they made assessment decisions. In particular, the following conceptions appeared to influence their summative assessment decisions and featured in descriptions of their practice: summative assessment as being improvement-focused (Section 7.2.1), the importance of student creativity (Section 7.2.2), and beliefs about positive engagement (Section 7.2.3).

7.2.1   Summative assessment as being improvement-focused

All beginner teachers in this study expressed a similar central conception of summative assessment: that it should contribute to an improvement in teachers' teaching and students' learning. They emphasised that summative assessment should not only be used as a final measurement for accountability purposes. Their improvement focus and responsive stance towards summative assessment meant the teachers felt it was very important to spend time considering the work and grades of their students, thinking about how they had taught them, and how this could be improved, and then taking actions based on their reflection. The amplifying and filtering effect of this improvement focus influenced a number of teacher assessment decisions and actions.

7.2.1.1   Improvement focused on improving teaching

All beginner teachers in this study believed that they should take actions to improve their teaching, based on their interpretations of summative assessment. Kate, for example, mentioned this early in the study as one of her underpinning beliefs, and then referred to it again at the end of her ITE programme, highlighting the unchanging nature of this belief over the period of the study:

> It's a way to gauge students on what they have learnt, what they know at the start, what they have learnt at the end, and then it can relate back to your teaching. Was my teaching, was I good enough? How could I do it differently? Looking at teaching as inquiry, was my learning … Was I teaching to the wrong learners? The right learners? Which ones did, and which ones didn't? Was my classroom management correct? Et cetera, et cetera … formulating assessment is to know what the students know or have learnt, and also as a way to direct back at yourself as to what you can change and what you can make sure. [Kate, I1]

> It's a way to gauge a student's whereabouts. So it would be gained throughout the course, are they getting what you are saying? Is it common in the classroom…? Are the students all gauging at the same level? Are there are a few students not getting it, a few students getting it? Some getting it better than others? So it is a gauge throughout the teaching and learning, and then it is the final assessment ... It is also a way of reflecting back as teachers … If the students are all getting question 1B wrong … Did I teach 1B or did I forget to? [Kate, I4]

All beginner teachers related stories which demonstrated how these beliefs amplified their actions. They were motivated to adjust or change their programme

to either re-teach or re-emphasise important content or skills where they identified areas of student misunderstanding, regardless of whether their schools promoted this view. They also discussed their decisions to speed up or move on to next topics when they were convinced that their students had a good grasp of what they had taught. In these ways, they focused on using summative assessment to improve learning outcomes for their students, as illustrated in these quotes:

> I was actually surprised by how much you can actually learn from the test. You see it and you see what is consistently being got wrong by students. You see where students are using incorrect terms to answer things, and so in that sense, just giving an assessment is not the end of the story. You've got a bit to reflect on, and a bit to think 'where to from here'? Essentially, now that we've got this sort of data, how do I then proceed from that? Rather than just 'they've done that, cool, move on'… That is a good amount of data to think about, to process, and try to incorporate it into the 'why?' for doing things. [Wiremu, I2]

> You have to have assessment to see how well the students are doing. … for qualifications, for next year and all the rest of it. But I think the main purpose is assessments for teachers to see where the main gaps are, where our teaching could improve, or where the students just need more learning. [Benjamin, I4]

Both of these quotes were from interviews that occurred after Wiremu and Benjamin had been working in examination oriented schools which encouraged very little reflection on summative assessment results; however, their beliefs about the importance of an improvement focus amplified their motivation to operate in this way. Where they felt the students required further teaching, five teachers discussed their need to examine their teaching approaches rather than reteach the topic in the same way. They felt the students' results showed them that the way they had taught the topic originally was not altogether effective:

> So it was good to be able to say, well this is a question the students didn't understand so next lesson I will be able to … Some didn't really get it so we'll go over it a little bit more … We might change their thought direction and try a different approach. [Kate, I2]

> So just being able to alter your teaching to direct what the content is to what their needs are, basically. Just finding ways to approach it from different angles. [Elisa, I4]

As part of their response to student misconceptions the teachers provided feedback and feed forward to clarify student ideas, often through discussion, going over test questions, showing students exemplars, and giving praise.

At times, beginner teachers talked about using the summative assessment results to modify their actions for the following year, rather than immediately. This was especially the case for end-of-year external NCEA examinations for senior students (Years 11–13), as illustrated in these quotes:

> I think it is something that if I was teaching I would look over the mock exams and the end of year exams, if you get to see them, and see where the struggles were. And then make that a priority next year to teach it better. Organise more personal reflection like, obviously that did not get taught as well as I could have, what can I change? What did work? What can I bring and to make it more interesting? Or more relevant for the students to learn? [Mary, I4]

The effect of teachers' improvement-focused beliefs was evident in their approach to NCEA. They understood the importance of providing quality feedback for school designed NCEA summative assessment tasks, as they deemed this would contribute to further learning for NCEA re-sits (for internally assessed standards) and for examinations (for externally assessed standards). In the following quote Elisa articulates the internal pressure she felt to make changes to her teaching in response to students' NCEA assessments:

> So if they haven't met the specific criteria then you have to… obviously, since you've assessed them… you have to figure out how to fix it. So what they haven't figured out. So if you've assessed yourself and what you've taught you have to be able to identify things to change as well. [Elisa, I1]

As a direct response to NCEA practise examinations and internal assessments, a number of teachers also talked about adding the teaching of assessment and examination techniques to their teaching programme.

7.2.1.2   *Improvement focused on improving student learning*
Beginner teacher beliefs about the importance of summative assessment being improvement focused meant they thought carefully about the relationship between student learning and summative assessment. Beliefs about summative assessment influenced their thinking about students' literacy, numeracy, their understanding

of the requirements of standards, as well as student motivation with these all being factors considered to influence final summative grades and marks. Elisa used elements of backward design when planning her teaching because she saw the links between gaps in students' science knowledge and skills and their more general literacy and numeracy skills:

> And so then we worked backwards from there. To work back to the very basics of stuff that wasn't even in the question but they have to be able to know in order to be able to answer the question. [Elisa, I2]

Generic skills were taught in order to help students improve their future science summative assessment results when these skills were noticed to be weak. Elisa's decision making was linked to her strong beliefs about assessment needing to be improvement focused, whether or not schools' science work plans included literacy as an outcome. Her beliefs amplified her assessment related decisions and actions in this regard.

7.2.2 The importance of student creativity

Another belief found to influence beginner teachers' assessment decisions focused on student creativity. This was raised by three teachers. Benjamin in particular felt strongly that schools that focused predominantly on summative assessment, particularly NCEA, tended to destroy student creativity because students became focused on passing assessments rather than on learning. In each interview he gave examples of this influence and he talked about wanting to improve students' general approaches to problem solving by encouraging creative responses. He was keen to encourage and evaluate students' creativity, rather than just focusing on meeting the requirements of standards. He talked about ways he would like to alter his summative assessment practice to mitigate against the negative consequences of a narrow assessment focus, and thus improve outcomes more generally:

> Their [senior students] imagination and creativity has just died down, and that's the same with these puzzles that I'm giving the students. So like the junior classes can … sometimes they are a lot better at answering the questions, and I think that is kind of like the way we've been assessed where everyone is focusing their attention, or all of their mental capacity into this area and just zoning it off ... I think it will narrow them down in their creativity, which I think is the big thing we

should be trying to build. Well, that's what I'm going to try and do as a teacher. [Benjamin, I4]

Ryan was another teacher who conceptualised learning science to include imagination and creativity. He wanted to encourage this in senior students, who he saw becoming too narrowly focused on NCEA results:

> Yes, that [NCEA] is too heavily emphasised and it encourages rote learning that I know doesn't work. I think that summative assessment is more effective when done in conjunction with something else, like an interactive activity or something that embeds it more. [Ryan, I4]

Both Benjamin and Ryan communicated a clear desire to adjust their assessment practice so that creativity was assessed alongside more traditional content.

The use of broader project-based summative assessment tasks which challenged students to draw from and apply a range of their science knowledge in order to find creative solutions to problems aligned with beginner teachers' beliefs about creativity. Through these sorts of assessment activities students were seen to have opportunities to draw on their learning across topics and synthesise what they had learnt, hence improving their general understanding of science and other subjects. Benjamin explained how he had done some research in the area of curriculum integration, particularly physics and mathematics, and he was looking for opportunities to assess aspects of both of these subjects together:

> But I would like to integrate it a lot more because I've done a lot of research … Like in physics, a student might be ace at maths, but the maths within physics he actually really struggles with. He actually doesn't know how to apply it. If I could apply the maths and physics and then I'll be able to assess as well how well they know that maths. [Benjamin, I4]

The use of integrated topics is evident in some internal NCEA assessment tasks which require senior students to draw upon a range of science knowledge. Susie talked animatedly about students learning through completing a research project on edible vaccines for NCEA (I4). This task required students to utilise and integrate science knowledge and scientific understandings that they had gained over a number of topics, so there was a creative element in the assessment. Susie talked about another unit as follows:

> In lessons you do little bites, and it all comes together in the summative. 'Oh, this all makes sense, and this relates to that bit over here'. Because you know science is taught in units but it relates so much to everything else, nothing is on its own ... That unit plan I just did, my theme was acid rain so it was kind of environmental in that sense. But it was reaction rates and concentrations, quite basic stuff. With a theme like that you can say, well scientists are doing this and they research theories in the outside world. It's not just what you did in science class. [Susie, I4]

In this quote Susie emphasises the benefits in teaching broad topics that integrate the learning from different topics, and in the artefacts she provided there was evidence of this. Similarly, she advocated integrated assessment tasks where this was possible, as she believed this encouraged creative thinking.

7.2.3   Beliefs about positive engagement

All beginner teachers in this study held strong beliefs about the need for students to engage positively with learning science, as they felt this led to improved learning. They believed that students should enjoy and be motivated by science learning as active participants, taking responsibility for their own learning. They emphasised the role of summative assessment as a motivator.

This next quote illustrates Elisa's strong view that once students have motivation in science, they will take more responsibility themselves for their learning.

> But science is different, not everyone enjoys it so they have got to have motivation to learn it…. I think it's when they can get to the point where they can choose to take it, then they start doing it for themselves. [Elisa, I3]

The teachers believed that students' enjoyment of science topics would be encouraged and enhanced if they enjoyed the associated summative assessment tasks and found success in them. Elisa talked about changing marking criteria to ensure junior students enjoyed success, thus decreasing the chances of them opting out of studying science in Year 11.

> All you want them to do is make them enjoy the subject. So it didn't really matter if I gave them a really easy test so they would get good grades and feel better after that …Yes, it was no big deal. I changed the marking scheme so it wasn't that hard to get Excellence. One of the kids was really good in class for me, and he got Excellence. He said to me, 'I told my mum I got Excellence and she didn't believe me, can I take my test home to show her please?' He was really

excited because he tried really hard in the test, and asked lots of questions during it ... It was a good payoff. [Elisa, I4]

This example shows how Elisa's beliefs about encouraging students to enjoy and positively engage in science acted as an amplifier, overruling her knowledge about quality in summative assessment, as she re-wrote the marking scheme and made it easier for students to score well. She did this in order to make sure the students felt good about science. There was a sense that, in general, students tended to view science as a difficult subject but teachers felt they could use summative assessment tasks to encourage engagement with science.

Ryan thought examinations were demotivating for a large number of students, particularly those with low literacy levels.

I don't think that formal exams are all that effective, so … And I also think they alienate a lot of kids from learning, so if it was up to me I'd probably be more oriented towards internal assessment, especially with younger kids. [Ryan, I1]

He believed examinations could contribute to the long term disengagement of students so as a result tried to avoid using examinations with his classes. In this case his beliefs acted as a filter, decreasing his motivation to use formal summative assessment tasks, especially examinations.

In this section the conceptions of beginner teachers about assessment needing to be improvement focused, the importance of student creativity and the necessity of students' positive engagement have been identified as acting as amplifiers and filters, affecting how they approach summative assessment and leading the teachers to make particular assessment decisions. These decisions were based on priorities resulting from these beliefs, and affected beginner teacher practice. In the following section I have chosen teachers' beliefs about ethical matters for special attention.

7.3 Ethical matters

A third amplifier and filter evident in beginner teachers' descriptions of summative assessment was their notions of what constituted ethical and fair practice in summative assessment. All teachers in the study raised the importance of ethically sound and fair treatment of students throughout the summative

assessment process. They talked about how their beliefs about the rights and needs of others affected their assessment-related decisions.

In this section elements of ethics in relation to summative assessment are presented in two sections: treating people with respect and care (Section 7.3.1), and honesty (Section 7.3.2).

7.3.1   Treating people with respect and care

In this study all beginner teachers talked about their sense of obligation to show respect and care for students in order to minimise harm. Wiremu's comment summed up their view of the need for ethical consideration with respect to assessment as follows:

> Because you can't have assessment without at least some types of ethical thoughts going on. I mean all things considered, what is best for the students, and for the profession, and for the school, and everything? There are a lot of people involved so you need to have quite an open mind, be open to what is fair from multiple perspectives. This brings ethics into it. [Wiremu, I3]

The teachers explained that they wrestled with the challenge of treating their diverse range of students equitably when summatively assessing within a school system. As discussed in Section 7.2.1, the teachers felt it was important to ensure that any summative assessment task that was used would provide some positive benefit for teachers and students, either benefitting their learning, contributing to their qualifications, or helping their confidence and self-concept as learners. In instances where they were given no choice in the summative assessment tasks they had to use, this meant the beginner teachers felt they needed to mitigate negative consequences for students whenever possible. The quote below from Benjamin highlights tensions they experienced in trying to meet their aspirations:

> Yes, because every student has strengths and weaknesses, so it is hard to make it completely fair for everyone. [Benjamin, I3]

Although their schools often insisted on common summative assessment tasks being used for all students, the beginner teachers ascertained that assessment tasks did not necessarily meet the needs of all students. They deemed the use of inappropriate assessment tasks unethical. It was evident that the beginner teachers were using their developing knowledge of curriculum, teaching, students, learning

and assessment even at the early stages of their careers to try to reconcile school requirements with what they saw as a fair and ethical practice. For example, Elisa explained how her AT used a common test for all Year 10 students, but she changed the test conditions so that weaker students would be able to refer to their books, enabling them to gain some confidence:

> He would give them the same test as the other Year 10s ... so I thought I'd change that because that sounded a bit ridiculous in my head, so I changed the test instead of giving them the same test. I gave them two periods to do it with their books [Elisa, I4]

All of the teachers recognised that one type of assessment activity was unlikely to suit all students, given the diversity in classrooms and they talked about the challenges of assessing struggling or disaffected students, students with disabilities, and students who had learning difficulties including low literacy levels. Four of the beginner teachers in this study discussed their feelings of obligation towards students who were already battling with self-doubt about their learning. Their commitment to learners meant they felt it was ethically correct to use summative assessment activities that allowed students the best opportunities to provide evidence of their knowledge and skill even if it was at a lower level, as a matter of respect and acknowledgement to the person. For example, for her Year 9 science class Susie raised the issue of giving students the best chance to show what they knew, and made decisions about allowing more flexibility in assessment design, including using oral assessment:

> Like if they don't like writing, don't give them a writing one … Like some oral presentations because some of the students were so adverse to writing… or giving them a chance of posters, because if you gave them a poster even if it was just brainstorming, it would be done. Whereas if you gave them writing it was a nightmare … Some kids can fully talk and tell you about it, but when it comes to putting pen to paper it can be a bit difficult. [Susie, I4]

In another instance, Elisa explained she had modified the school-wide Year 10 science assessment task so that it allowed students who were working at a lower curriculum level to demonstrate what they knew; for example, by adding more visual clues:

> Well like with our assessment for the Year 10s we had pictures that they could label, and diagrams that they had to answer little bits on,

and then there was the question where they had to compare and contrast and write a paragraph. So the ones that knew the pictures could do that …. At least they could show that they know something. [Elisa, I2]

By making the decision to add questions which included diagrams and more scaffolded tasks, Elisa gave the academically weaker students better opportunities to show what they had learnt and gave them questions they could get correct. Her comments indicated that she believed this was the ethical way to treat students. She respected them as people and did not ignore their progress in the science class, albeit that their final achievement might be at a lower level than the majority of the students. She felt it was important for the students' self-esteem that they were able to score reasonable marks in their assessments, minimising harm.

Similarly, Benjamin commented on the need to use a wide range of assessment activities or modes of communication as a way of allowing all students a chance to show what they knew and could do. He argued that for students with literacy difficulties, written assessment tasks were not appropriate, and he regarded this an ethical issue:

Some students aren't very good at writing what's in their head down on paper but they can come up with some really, really good ideas. So ways to assess that … But they don't really show it when you're talking to them, some kids are a bit quieter. Some kids just don't show it. So I suppose you've got to assess kids in lots of different ways. 'Cos one on one assessment doesn't suit some kids but it does suit others. Written assessment like that doesn't work for everyone either. [Benjamin, I1]

These beliefs acted as amplifiers and filters that affected their decision making about the focus for student learning and assessment. This meant at times they planned alternative assessment tasks, and assessed in ways that contradicted their schools' policy. The approach of using a different range of assessment tools meant extra work for the teachers, but they valued their students and saw this practice as consistent with the ethical treatment of students. They also thought this provided more valid and reliable assessment data, as well as minimising harm for students.

Another example of how teachers might treat students ethically for summative assessment was the acknowledgement of students' first languages. This was discussed with respect to assessment by Wiremu, who felt that it was his ethical duty to give Māori students the opportunity to be assessed in te reo Māori if they so wished, even when working in an English-medium school. This was despite the school he was working in not being supportive of this stance. He gave an example of a time when he trialled an oral assessment of science in te reo Māori but his students found the science terminology and concepts difficult to work with in te reo Māori, so they ended up reverting to English. Given that the decision to revert to English was the students', Wiremu saw this as an important element of his respectful treatment of the Māori students (I, 4).

When investigating teachers' conceptions of the ethic of care, some interesting and somewhat conflicting findings emerged. As discussed in 6.2.2, at times teachers altered marking schemes, making the criteria 'softer' to enable students to gain better marks in the belief that this would keep them engaged in science and be better for their self-esteem, so there was an element of an ethic of care in their practices. However, these sorts of decisions could be seen as being unethical, undermining the schools' assessment system. As a contrast to this, two teachers interpreted their ethic of care as trying to make sure students achieved the highest grades they were capable of getting in NCEA. That is, the teachers cared about maximising the students' academic achievement. They did this by increasing the difficulty of assessment tasks over time in the belief that by being tougher when applying criteria this would keep students from getting inflated ideas about their progress or capability. They felt this would keep the students learning and striving for higher grades, as is illustrated in this quote referring to practise examinations:

> You don't want to be too nice and give them … They get a big head from getting that question right, that's kind of what we got told. Don't give them an Excellence unless it is a perfect answer, which is, well it's not going to be perfect. None of them got Excellences, which was a good thing. [Elisa, I2]

> But I realise that it is better to mark them harder in the [practise assessments], so they can see what they need to work on, and I kind of came to that realisation like halfway through ... so in the long run it is not going to be helpful … So 'be nice' by marking them hard? [Bella, I4]

Both Elisa and Bella believed that not getting good grades would spur students on to work harder and end up with better grades. However, this sort of caring for academic excellence could also be perceived as unethical treatment of students, given the apparent lack of honesty and transparency. Nevertheless, these two teachers' assessment practice appeared to be in part shaped by their beliefs regarding their care for helping students gain high student achievement.

All beginner teachers in the study recognised their position of power within the relationships they developed at school, particularly their power over students. Ethical behaviour to them included the prevention of abuse of power. They acknowledged their responsibility to manage power relations ethically. Historically, teachers have made all the decisions regarding summative assessment, and students have had no input into how and when they are assessed. This status quo was challenged by teachers in this study, as many of them felt it was ethically beneficial to share power and enable students to have input into the summative assessment process. Five of the beginner teachers described ways they decreased the power differential by allowing students to contribute to summative assessment practices, in particular by negotiating the task-type of summative assessment for given topics. Examples given by teachers included allowing students to choose whether to present information orally or in written form, enabling students to choose contexts for their NCEA biology assessments, co-constructing marking rubrics for laboratory reports with students, and letting students choose whether the assessment was open book or closed book. Power sharing extended to teachers allowing students within the same class to choose different modes through which to demonstrate their learning on the basis that this would be "more fair".

> So I think that summative assessment is … because it is the same for everyone, but maybe a better assessment might be to let students choose whether they want to give a written essay or give an oral presentation. I think that would be more fair. [Benjamin, I4]

Four teachers described the lack of flexibility some schools permitted in co-constructing assessment tasks as poor practice, as it did not fully support the rights of students to have a fair opportunity to show teachers what they knew and could do. Even if changes such as adjusting time frames in consultation with the

students appeared minimal, the teachers felt they were valuable, as they allowed for a shift in the balance of power in the classroom. Three teachers felt they did not have the authority to share power with student within the summative assessment process. All the teachers conceded that final judgements of assessments contributing to high stakes assessments remained with them.

Overall, teachers in this study clearly articulated a belief in students being treated in an ethical manner with a focus on fairness and respect. They admitted that this sometimes caused challenges for them and competing priorities sometimes led them to change assessment conditions, particularly for Year 9 and 10 students. However, ethical consideration was an underlying principle for the beginner teachers and did affect what they paid attention to. At times they chose to assess in ways that contradicted their schools' standard practice or policies, illustrating the amplifying/filtering effect of their beliefs on their decision making.

7.3.2   Honesty

All beginner teachers in this study identified as a strong belief their need to be honest in their dealings with students, parents and other stakeholders, and with themselves. They felt that their decisions regarding summative assessment needed to be open to scrutiny, but as described above they admitted that competing pressures sometimes forced them to compromise on some aspects at some level.

As inexperienced assessors, the beginner teachers were particularly concerned to ensure their marking and grading represented an honest appraisal of students' work. They worried that their judgements might be biased. For example, Benjamin felt that marking anonymous student work might mitigate some of the temptation to bias their marking:

> Like I think for summative assessment you should try not to look at whose paper it is, because I think some teachers, some teachers were influenced by who the student was, and that is how they were marked. Like if a teacher didn't like a student, they generally got poorer marks, if the work was quite similar quality. [Benjamin, I4]

Susie was happy when she saw marking being shared amongst the teachers in her school, as she felt it increased the chance of grades being an honest appraisal of student work because it reduced the potential for individuals to bias the marking.

Grading through the use of the criteria provided in the standards was seen as a way to ensure the honesty of teacher judgements:

> It's like everything has to be based on a criteria and justified somehow. [Wiremu, I3]

> They seemed to share it amongst the classes they were doing the assessments with, so they would mark half a class each so there wouldn't be bias and they weren't just marking their own kids. It ended up being quite a nice way to do it, and the classes were relatively balanced. [Susie, I4]

Wiremu commented on the use of criteria to justify grades, and these are also the basis on which moderation works.

However, Bella in particular found it difficult to give students failing marks or low grades:

> Although decisions came into it, I hated marking to be honest. Like if they were so close [to passing], I just feel mean … I don't even know the students but I still feel mean … well I think, because I should have failed a couple of them, and I know that because they hadn't related [their research] to other sources, but they had Achieved with Excellence in other areas [of the task] so I couldn't justify failing them. So I 'Achieved' them. [Bella, I4]

As illustrated in this example, she made a conscious decision to be less than honest by awarding the Achieved grade when the students had not shown evidence of meeting all of the criteria when she marked a class set of practise assessments. She wanted the students to pass the standard and she felt mean if this did not happen although she felt very conflicted about her decision to pass students. This example highlights Bella's seemingly unethical practice, as she should have judged the student work against the standard but her emotional concern for the students overrode her knowledge of quality marking and she did not apply the marking schedule accurately.

Discussing another perspective on honesty for summative assessment, two of the teachers raised issues of content validity of tasks, and in particular the need for assessment tasks to match specific learning objectives (SLOs). This is important if the results are to honestly represent student achievement. In some instances

school-designed tests were deemed not to match the learning students were experiencing:

> I was given the SLOs in the scheme, so I knew what was needed to be covered, and so I kind of covered what was asked of me. Like the SLO about understanding curved mirrors and their uses, so you cover that. And then when it comes into the exam.... Like there were questions about lenses and prisms which were nowhere in the work we did, or examples that didn't reflect any of the information we covered. [Susie, I2]

In this instance, Susie found that the assessment task did not really match the learning outcomes she was working towards, which was unfair for the students. Grades or results from the assessment in this case did not honestly represent the learning achieved by the students. Susie's natural reaction to this was to change the tasks, even though she had been told to use what she was given.

Adding to this, three teachers raised the issue of students deserving a completely honest appraisal of their progress, even when the appraisal was negative. Wiremu spoke of students needing to be treated as scholars and needing to be able to accept honest results even when they were not positive.

> I mean you can't just say I gave him an Achieved, there has got to be a reason behind it. Is this fair? "I thought he did enough effort in class, it wasn't enough for a Merit, so I'll just chuck him an Achieved because he hasn't failed?" No. [Wiremu, I3]

Wiremu's sense of honesty meant he was more likely to bluntly tell students about their weaknesses than other beginner teachers.

Within the NCEA system, the built-in processes of moderation, both in-school moderation and the national external moderation system, has been designed to ensure national consistency in assessment. Beginner teachers in this study believed strongly in this system but were critical of less than honest practices they saw in schools. In particular, one teacher described the manipulation of the student work that was sent for external moderation. This school purposefully selected student work instead of using the official random selection process. This was seen as dishonest and unethical behaviour as it meant that the system for quality assurance was being circumvented. In speaking about the selection of student work for moderation of marking Benjamin commented,

> But even talking about that [moderation], with fairness, like the teacher is trying to filter through to find a solid Achieved, a solid Merit, and a solid Excellence, rather than a borderline one just to make sure they are trying to get right in the middle. So you don't just pick anyone's out. So I don't think it's a fair moderation. [Benjamin, I withheld so the school cannot be identified]

All teachers in this study agreed that it was their responsibility to ensure that the processes put in place to assure quality of the NCEA assessments were adhered to.

All teachers in this study were concerned about the temptation to teach to the test in order to help their students achieve better results in summative assessments. They felt it was dishonest for teachers to teach to the test as not only was it cheating the system but it also disenfranchised the students as competent learners. They all gave examples of experiences where they had seen teachers teaching to the test, especially for NCEA internal ASs assessments. The teachers were united in their view that teaching to the test undermined the summative assessment system, potentially producing less than honest reflections of students' learning. They understood the need to prepare students for summative assessments but felt a number of teachers went beyond what was ethical by teaching to the test. As Kate observed:

> It's a little bit, you can get a bit negative with some of your associates. You look at them and they say, "Well this is going to be in the test so this is going to be in the exam, this is going to be what you are going to get marked on, this is it. You do bad in this or good in the topic because this is what is going to be causing you to pass or fail." I think the way a teacher reacts around that can be quite negative to the students, or to the teaching in itself. [Kate, I3]

Their strong views amplified their commitment to teaching a broader curriculum.

Making sure that students were being honest in the work they submitted for internal summative assessment was an obvious ethical requirement in summative assessment, and one that was recognised by the beginner teachers in this study. Given the nature of many of the NCEA internal assessments, which were completed over a period of days or weeks, the teachers understood their need to ensure the authenticity of students' work. They discussed their responsibilities to assure that the work marked was authentic and noted that they had learnt a number of strategies to assist with this during their practicum experiences. These

are illustrated in the following comment from Kate, when she was reflecting on her decision to get students to write up their internal assessment report in class instead of allowing them to do it in their own time:

> And it [the need for authenticity] would be the same in science, make sure they are doing the right method. In science, like, they'd still have to be able to write a report, write out their methods. But at least if you are seeing what they are doing then they are not going to turn around and write down what the kids next door did. [Kate, I2]

In conclusion, all of the teachers in this study raised issues of an ethical nature and commented on the importance of fairness for all students. The teachers struggled at times with ethical dilemmas when their beliefs about ethical practices were challenged by what they were asked to do. In this way the ethical beliefs the teachers held could be seen to affect their summative assessment decision making. Beliefs relating to ethics amplified or filtered their options and preferred actions. A fourth amplifier and filter somewhat related to ethics and found to affect the teachers' assessment practice was teacher emotional response. This is discussed in the following section.

7.4   Teacher emotional response

Emotion acted as an amplifier and filter for summative assessment practice and decision making for the teachers in this study. All the beginner teachers talked about their emotions and feelings when they reflected on their summative assessment practice. Their emotional responses were seen to be influential on their classroom assessment decisions and on their confidence when they were being assessed by others and when they were assessing others. The teachers discussed their emotions from the perspectives of assessing as teachers (Section 7.4.1), and being assessed as teachers (Section 7.4.2).

7.4.1   Assessing as teachers

Five of the teachers talked about their struggles with emotion as they tried to make good assessment decisions. These teachers mentioned they felt scared and worried about the effects of the grades they assigned had on the students themselves. They sometimes felt they were being unkind and uncaring if they gave students low grades.

Elisa made this point when she discussed her experience of assessing an academically weak Year 10 science class. She wanted to measure their learning progress against the class SLOs but felt strongly that she needed to help the students feel positive about learning science. For the assessment task she gave the class 'hints' on the board and she adjusted the marking criteria in order for her class to gain higher grades:

> I did all the criteria, and on the board I put all of the things they needed to do like "you need to have the units", "you've got to read the question so you'll know what units to use" ... It was a good payoff. [Elisa, I4]

She talked about feeling guilty about doing this, but felt that in the end the positive responses received from the students justified the lack of rigour in her assessment process and marking. Again, this demonstrates how, in some summative assessment decisions, emotion overrode expected practice. In this example Elisa talked about her decision making as a process in which she had to weigh up a range of factors, including emotional responses, in order to get the best payoff for the context.

When the beginner teachers interacted with their ATs, mentor teachers and other teachers about summative assessment issues, particularly grading NCEA tasks, this had a considerable influence on their levels of confidence, and sometimes resulted in them feeling very stressed. When their ATs were confident in the NCEA system, communicating that it was trustworthy, teachers also felt more confident and able to engage in the system. However, when the ATs themselves were not happy working with the NCEA system this led to decreased levels of confidence and further uncertainty for these novice teachers, as illustrated in this quote by Bella:

> There was a lot of chatter in the science workroom between the two biology teachers, and the moderation, they were not 'confused', but were debating whether the student should get an Achieved or Merit because it was so vague … well it was quite negative, quite negative about the standards really … when it came to marking I felt there was a lot of negative vibes, not knowing how to rank the students because of its vagueness. [Bella, I2]

In this comment Bella described her colleagues as being uncertain about their marking, and as a result she became less confident about her ability to mark against the achievement standard. The negative emotional side of such interactions acted as a filter, contributing to a lack of confidence in how she marked students' work. She explained that she struggled with indecision and battled with a reluctance to award a grade, for fear of getting it wrong.

The next example from Kate illustrates the link between teacher beliefs and emotions. On a practicum Kate found the practices of her AT difficult to reconcile with her own beliefs when she was asked to run a particular assessment. Her resulting emotional state of confusion and worry meant she found it hard to make decisions at times:

> You can get a bit negative with some of your associates. You look at them and they say, well this is going to be in the test so this is going to be in the exam, … I didn't really think it was a good gauge of where the students were at … and I felt a bit jaded as well, and the fact that she had told me to ask these questions, like "there is a set of questions, ask these". [Kate, I4]

In these examples interactions with ATs negatively influenced the teachers' emotional states and levels of confidence when making assessment decisions. Bella struggled to mark student work, and Kate felt unable to make decisions about assessment questions. When these sorts of experiences were recalled by the teachers in this study, they mentioned feelings of confusion, worry and doubt.

On the other hand, the actions of some ATs contributed to building teacher confidence by providing positive feedback or showing they had faith in them, resulting in the teachers feeling calm and capable. For example, Elisa's level of worry and concern dropped when she felt trusted by her AT:

> I did marking for Year 11s, for their practice test, which … My associate was just kind of like … "I trust you", so that's fine. She didn't really go over my marking, but I did have lots of questions for her at the beginning, and we went through all of those. So she was confident with how I was doing it. [Elisa, I2]

This positive emotional experience led Elisa to being confident in her summative assessment decisions; she did not feel she needed to check for approval from her

colleagues, as the positive emotions amplified her levels of confidence in her ability to be able to make assessment decisions.

In summary, five of the eight teachers expressed emotional responses about assessing as teachers. These emotions were seen to affect their confidence and focus when making summative assessment decisions.

7.4.2 Being assessed as teachers

Beginner teachers felt that the results their students achieved in summative assessments would be used by other teachers to judge their teaching competency. The teachers talked about this as being quite draining emotionally as they dealt with feelings of vulnerability and uncertainty. Wiremu and Mary articulated these fears, and the self-doubt that resulted when their students did not do well in summative assessments:

> I guess that is the biggest fear, it's just, oh man, imagine if I did this whole test and they all failed. I'd suck as a teacher … I felt it was on me if they didn't pass, like it was my fault because I hadn't taught them correctly. [Mary, I2]

> I am kind of nervous about what my judgement is going to be. Am I going to be judged based on my judgement of others? And so these things all go through your head. [Wiremu, I4]

Mary observed that she was not the only teacher who was concerned about being judged by other teachers. She noticed that in school other teachers were cautious about comparing classes:

> I think with what I saw at [a school's name], everyone is very … they don't want to see their class not doing well when compared to someone else's class. And so you do see that kind of struggle, like how well is your class doing? Because it is doing the same as mine so we're alright then. [Mary, I4]

This quote highlights the nervousness teachers feel in being judged or compared to other teachers based on their students' performance.

These teacher responses indicated they held a deep-seated fear of being found lacking as a teacher. Over time the fear of being found lacking as a teacher by other teachers did not seem to diminish. Later in the study, Wiremu realised that he was making judgements about one of his ATs based on how well the AT's

246

students performed. Wiremu explained that if he was making judgements about other teachers, they in turn would be making judgements about him, and this made him feel uncomfortable and vulnerable. As a result of the emotions evoked about being judged by other teachers, the beginner teachers lacked confidence to share summative assessment data and to speak about it with other staff.

The beginner teachers also assessed themselves based on the summative assessment results generated by students and took these results very personally. Among others, Kate indicated the responsibility she felt for her students, and how this scared her:

> That's what I'm not a little bit scared of but I don't want to feel that I have let the kids down. [Kate, I4]

Similarly, others felt quite a personal burden when it came to the self-assessment of their senior classes, so that if any students failed this meant they were poor teachers in their own eyes:

> I felt it was on me if they didn't pass, like it was my fault because I hadn't taught them correctly. [Mary, I2]

> Well I had a bit of taste for it and I didn't realise I would take it so personally, all the bad results. I was just like, "I remember teaching you this". I remembered having conversations. [Wiremu, I2]

Later in the year Mary reflected on the likely causes of their students' results, and by then had become more reconciled to the idea that students' results were not solely her responsibility:

> But just seeing that just because your students don't do well doesn't mean you're not a good teacher. It just depends on the student or how the exam was worded … There are all these excuses and it is good not to rely on excuses but teachers do their best. [Mary, I4]

The teachers in this study struggled to find the balance between reacting emotionally to students' assessment results because of a feeling of personal responsibility, and stepping far enough away to realise that the students themselves were responsible for their own learning. However, by Interview 5 four of the teachers seemed to have struck a balance, as is illustrated in Wiremu's reflection:

Because I have two Year 9s and I have two Year 7–8s [classes], and we're teaching the same topics for usually the same length of time, so it's really great to see how the results can vary. And it has helped me recognise that I shouldn't take the results so personally because I can't control them. They are not a direct reflection of me. Yes, they probably reflect some [of] my teaching of course, but that grade is not my work. [Wiremu, I5]

Wiremu had moved from taking full responsibility for students' grades which generated intense emotional responses, to attributing the grades to the students themselves, at least in part. This reduced the emotional load he felt while marking work.

Susie and Bella felt more relaxed and happy about their teaching when their students gained good results in summative assessments. They saw these results as confirming their expertise as teachers. They talked about the positive thrill they got from seeing their students doing well in summative assessment tasks, and they felt this reflected well on them as teachers:

Like, okay, you're testing the kids but you're kind of testing yourself at the same time. It's quite amazing. [Susie, I2]

My associate said no-one has failed yet, so I must have been doing something right. [Bella, I4]

So if the classes did well, the teachers tended to be very happy and excited, but if the classes did poorly, the teachers felt very discouraged and inadequate.

Benjamin was the only teacher in the study who did not express the same nervousness about being judged by himself or other teachers based on his students' summative assessment results. He did see the students' results as being a reflection of his teaching.

But now it is also a reflection on the teacher, and what they need to change. You can assess what different approaches teachers are taking. Assessment is now, I think it is pretty 50/50 the student and the teacher, because the assessment is about the teacher as well, and what is actually working. Whether teaching techniques are working for them. [Benjamin, I1]

However, he saw this judgement as being a necessary part of improving teacher practice.

> Yes, I think it is important to have the teachers assessed as well as the students I think, for the benefit of the students. [Benjamin, I2]

Benjamin appeared to approach his work more clinically, and possibly because of his previous work experience he welcomed critique without expressing an obvious emotional response.

In summary, the beginner teachers in this study showed evidence of being emotionally involved in the process of summative assessment. They communicated the real fears and apprehension that they felt, and then in some cases their affirmation when they looked at their students' results. They were often nervous and fearful of the judgements others might make about them as teachers, based on their classes' results. There was little evidence that they talked about these fears with anyone. The beginner teachers indicated that emotion tended to cause them to be more tentative in making summative assessment decisions and was a source of worry for them.

7.5   Summary

The findings in this study indicated that a number of factors acted as amplifiers and filters, influencing teachers' summative assessment decision making and practice. Each teacher described the ways in which sociocultural context, their conceptions of assessment, ethical matters and emotional responses influenced their approach to assessment and summative assessment decisions. At times the factors described in this chapter acted as amplifiers, bringing certain aspects of knowledge or priorities to the forefront in decision making, and at other times these factors acted as filters, reducing or decreasing the attention teachers accorded to certain aspects of practice. The teachers said that competing factors sometimes caused conflict or indecision for them, but at other times confirmed their next steps of action. At times these factors generated tension for individual teachers, and their reasons for making assessment decisions were complex and not necessarily expected. It was clear, however, that these amplifiers and filters influenced teachers' summative assessment decision making.

Chapter 8: Discussion

This study investigated the characteristics and development of summative assessment literacy for eight secondary beginner science teachers. The study's aim was to build knowledge about the characteristics of summative assessment literacy and to gain understandings of the ways and means of summative assessment literacy development. Since this required an in-depth investigation, a qualitative interpretive approach was used, utilising data gathered over time. As explained in Chapter 4, thematic analysis and the application of the SALRubric enabled an interrogation of the characteristics of summative assessment literacy and its development.

This chapter discusses the findings of this study under three main headings which directly respond to the research questions. The first section (Section 8.1) focuses on characteristics of summative assessment literacy, the second section (Section 8.2) discusses the development of summative assessment literacy over time, and the final section (Section 8.3) discusses what beginner teachers identified as contributors to their development of summative assessment literacy.

8.1   Characteristics of summative assessment literacy

This section responds to the first research question, "What are the characteristics of summative assessment literacy as perceived by beginner secondary science teachers?" The characteristics are discussed by firstly focusing on the knowledges teachers view as being encompassed in summative assessment (Section 8.1.1), and secondly by considering the personal and contextual factors that act as amplifiers and/or filters for what beginner teachers choose to implement in practice (Section 8.1.2).

8.1.1   Knowledges of summative assessment

Beginner teachers develop considerable understanding of the knowledges involved in summative assessment early in their careers. The dimensions of knowledge have been categorised for the purposes of this thesis as knowledge about the purposes of assessment, knowledge of what to assess, knowledge of summative strategies and task design, knowledge of summative assessment

interpretation, and knowledge of quality in summative assessment. Although treated separately in this section, knowledge of summative assessment is an amalgam, as a blend or mix of these various knowledges.

8.1.1.1 *Knowledge of purposes of summative assessment*
Beginner teachers were able to identify and talk in detail about a range of purposes for summative assessment with some nuance.

A prominent finding was that all beginner teachers in this study considered that summative assessment enabled teachers to gauge students' learning so that they could use the information to inform their pedagogical decisions (see Section 5.1.1). They reported that summative assessment could inform teachers about how they might best teach by adjusting their teaching, re-teaching topics and/or planning for future changes. Given the nature of education policy and practice in New Zealand, which has promoted formative assessment for many years, including the formative use of summative evidence (see Section 2.2.1), this finding is not altogether surprising. It is also consistent with New Zealand and USA studies which have found that teachers do make changes to their instruction and assessment practices as a result of student assessment grades (McMillan, 2005; Siegel & Wissehr, 2011; Timperley & Parr, 2005). In science education there is a strongly constructivist tradition, whose principles are reflected in formative assessment (Black & Wiliam, 2009), to which all the beginner teachers in this study subscribed. The models of learning that teachers hold have been found to influence assessment task design, practice and use in other contexts (Abell & Siegel 2011; Pelligrino et al., 2001; Penuel & Shepard, 2017; Shepard, 2000), so it could be expected that teachers in this study view summative assessment as having a formative purpose as well as a summative purpose.

Interestingly, the beginner teachers in this study also spoke of the formative use of aggregated summative data although they appeared less knowledgeable about this use. Their reported lack of opportunity to learn about analysis and the use of aggregated summative assessment data (see Section 5.1.1) reiterates the claims made by other researchers who have found that teachers are often unskilled in the use and analysis of aggregated data (Datnow, Park, & Wohlsletter, 2007; Earl & Katz, 2006; Mandinach & Gummer, 2016; Pierce & Chick, 2010; Schildkamp &

Kuiper, 2010). Studies have shown that this is an area which requires explicit and ongoing professional development (Mandinach & Gummer, 2013; Schildkamp et al., 2013; Unger, 2013).

Summative assessment is used within schools for reporting as well as for class allocations and streaming. The beginner teachers identified these functions as purposes for summative assessment. They considered they knew enough to confidently comply with school reporting requirements that were expected; however, they were less sure about streaming. Streaming is used in most New Zealand secondary schools and involves students being placed in ability groupings or classes (Hornby & Witte, 2014; Hornby, Witte, & Mitchell, 2011). Beginner teachers knew that streaming directly linked to students' performance in summative assessment (see Section 5.1.2) but had less understanding about the technical processes involved. This may be an artefact of when they were exposed to work in schools, given that practicum periods did not coincide with the times of the year when classes were being organised.

Summative assessment is typically associated with the award of exit qualifications. Unsurprisingly, beginner teachers recognised and understood this as a purpose for summative assessment (see Section 5.1.4). Over time they reported a developing understanding of standards-based assessment, and confidence to work within this system (see Section 6.1.2.1). Nonetheless, by the end of this study around one third of the eight beginner teachers were not confident in their understanding of the requirements for quality NCEA assessment. Task design was identified as being particularly challenging. These findings mirror those from studies by Hipkins (2013), Hume and Coll (2010), and Anthony and Kane (2008), and signal that more needs to be done to prepare teachers for their role in this high stakes assessment.

Beginner teachers spoke of the learning purpose of summative assessment, in which project-based summative assessment activities caused students to draw together knowledge from a number of science domains and synthesise what they had learnt (see Section 5.1.6). Using summative assessment to provide students with opportunities to learn provides a counterpoint to the finding that it can be used to enhance teaching. The use of assessment tools as a means of supporting

learning has been a topic of interest for a considerable time and research shows that assessment can be used successfully in this way (Carless, 2007; Earl, 2013; Torrance, 2007). This use of summative assessment as learning is a useful strategy for the New Zealand context, given the extreme time pressure under which some teachers of NCEA classes find themselves, being beneficial for both students' learning and by taking the focus of the assessment away from it being purely for accountability and reporting.

The beginner teachers in this study also considered summative assessment as a way to motivate student learning and prompt more productive behaviour as a means to spur students to be more focused in class and to work harder (see Section 5.1.5). They implied the relationship between motivation and assessment was one of simple cause and effect and appeared to be working within a naïve belief that the way they marked or graded NCEA assessments would directly affect student motivation and engagement. In contrast, Crooks (1988) described the relationship between motivation and assessment as complex, mutual, and mediated by student self-perception. More recent reviews have added a more nuanced explanation, including an emphasis on the agency of students and their self-regulation as they are involved in assessment (Black & Wiliam, 1998b; Brookhart, 2013c; Harlen & Deakin Crick, 2003). The work in New Zealand secondary schools by Meyer et al. (2007, 2009) confirms the complexity of students' motivation orientations with respect to NCEA showing that a cause and effect relationship is not sufficient to explain student motivation. This being the case, it would be prudent that beginner teachers are made more aware of the complexities involved for student motivation.

8.1.1.2 *Knowledge of what to assess*
Beginner teachers had some, but limited, knowledge of what to assess. On practicum they were all required to teach aspects each of biology, chemistry, physics, earth sciences and astronomy, which meant they had to teach and assess science content they had not necessarily learnt at university. As would be expected, they were positive and confident about knowing what to assess for familiar topics, and their comments indicated they had less confidence for less familiar topics (see Section 5.2.1). Researchers have found that science teachers in many contexts may have limited science content knowledge as well as science

254

misconceptions (Hope & Townsend, 1983; Kind, 2014; Kruse & Roehrig, 2005; Loughran, Mulhall, & Berry, 2008; Yilmaz-Tuzun, 2007). Accordingly, teachers working outside of their subject specialisation have been found to act like novices with a less sophisticated approach to teaching because they have their own misconceptions or gaps in knowledge. In contrast, those working in their specialisation are working in an area of strength (Arzi & White, 2008; Hashweh, 1987; Houseal, Abd-El-Khalick, & Destefano, 2014; Luft, Hill, Weeks, Raven, & Nixon, 2013). This being to case, it is understandable that the beginner teachers in this study were less confident in deciding what to assess for less familiar curriculum content, but it does raise a concern regarding teacher preparedness (Cameron & Baker, 2004).

A further finding regarding what to assess was that beginner teachers knew about the importance of assessing skills that were not science-specific, such as graphing, creativity, curiosity and questioning. Skills such as these are important for science but are not limited to being taught in science classes and are often not explicitly assessed or reported on (Abrahams, Reiss, & Sharpe, 2013; Lai et al., 2016). This finding indicated therefore that beginner teachers understood the importance of aligning what they assessed with the breadth of what they taught, rather than restricting their assessment down to a very narrow focus on science content.

Two aspects were surprisingly absent from the teachers' knowledge of what to assess. Firstly, teachers did not comment on assessment of three important areas of the New Zealand Curriculum: the Nature of Science (NOS), key competencies, and the inquiry process. International research continues to find that teachers themselves do not hold adequate conceptions of NOS (Abd-El-Khalick, 2014; Hipkins, Barker, & Bolstad, 2005; Lederman, 2007) and this being the case, they may be less likely to assess NOS. Similarly, it could be argued that beginner teachers may not have sufficient understanding and experience to explicitly think about how to assess key competencies and the inquiry process.

Secondly, beginner teachers did not appear to consider the importance of assessing central or key science concepts required as a basis for building other science knowledge. Instead, they treated all science knowledge and skills as more or less equally important for assessment. The teachers learnt about using CoRe

design (Loughran et al., 2006), and from this work they identified key ideas central to teaching particular science topics. Hume and Berry (2010) found that CoRes were helpful but challenging for preservice teachers as their lack of classroom experience and experimentation limited their usefulness. Somewhat disappointing then was the finding that although specific important science knowledge and skills were identified by beginner teachers, this did not lead to them prioritising which were most important to summatively assess.

8.1.1.3   *Knowledge of summative strategies and task design*

Beginner teacher knowledge of summative assessment strategies was broad and they were able to describe a wide range of assessment types in general terms (see Section 5.3). Beginner teachers expressed the desire to use a variety of assessment types, particularly tasks that were more authentic and that integrated the assessment of a number of learning objectives. They were motivated to provide authentic tasks that linked with twenty-first century skills in a manner consistent with international trends towards teachers using authentic assessment (Koh, 2014; Scardamalia et al., 2012).

However, all the beginner teachers were concerned about their capacity to design summative assessment tasks. They felt they needed more knowledge of generic summative assessment task design, as well as context-specific task design (in their case, NCEA task design). In this study the first practicum was noticeable as a potential catalyst in the beginner teachers' development of knowledge of strategies and task design (see Section 6.1.1.4), but development after this appeared very gradual across the group. Susie was an exception to this, as she showed exceptional development. The beginner teachers' lack of knowledge and confidence in task design echoes findings from earlier New Zealand studies (Anthony & Kane, 2008; Smith et al., 2013) as well as international studies in which the need for teacher instruction in assessment task design has been highlighted (Brookhart, 2013b; DeLuca & Klinger, 2010; Lyon, 2013b; Popham, 2011; Yilmaz-Tuzun, 2007).

Findings indicate that a concern with equity and fairness for task is an aspect of summative assessment knowledge. The beginner teachers talked about ethics and fairness when they discussed summative assessment task design. They were very

aware of the need to design tasks that allowed all students the best opportunity to produce evidence of their learning. Similar to other studies focusing on ethical and fair assessment, they added concerns about the potential roadblocks caused by literacy and other demands in the way science tasks are designed (Bullough, 2011; Green et al., 2007).

8.1.1.4 *Knowledge of summative assessment interpretation*

Another element contributing to the amalgam of knowledges of summative assessment is the knowledge of summative assessment interpretation. Findings indicate that beginner teachers had a limited but growing knowledge of how to mark and make judgements against standards and to interpret the meaning of grades (see Section 5.4). This finding is in line with commentary about the New Zealand qualifications system by Black (2001), who remarked that it takes time for teachers to build expertise in making judgements. More recently in their report on moderation and teacher learning in New Zealand, Hipkins and Robinson (2011) commented that teachers build expertise over time and through the opportunities for professional learning that social moderation affords. The beginner teachers felt they did gain a good understanding of the assessment criteria for the specific NCEA assessments they had to mark in school, as they often marked in conjunction with other teachers, which they found very helpful. This finding lends support to the contention that assessment can be usefully viewed from a sociocultural perspective (Klenowski & Wyatt-Smith, 2014), and that beginner teachers' interactions enable them as a community to develop and articulate their understandings of standards as they draw on both their own tacit and explicit knowledge as well as the tacit and explicit knowledge of the group.

In contrast, the summative assessment tasks that the beginner teachers used for non-NCEA tasks and junior (Years 9–10) students were much more variable, and this made interpretation more challenging, sometimes leading to confusion or marking inconsistencies. Moderation only occurred for NCEA assessments, so for Year 9–10 classes the beginner teachers did not gain the benefits of a shared understanding of assessment expectations and instead marked work independently. The diversity in student population and contexts further challenged teachers to make judgements of a wide range of evidence of student learning. For example, beginner teachers faced dilemmas over the grading of work

from students with learning difficulties (e.g., low literacy levels) or behavioural challenges, and in some cases they responded by making changes to assessments; for example, by adjusting marking criteria or marks. This finding resonates with the research by McMillan (2001) and Randall and Engelhard (2010), who argue that teachers use a combination of factors when determining grades rather than basing grades on achievement alone. Similarly other studies have found that factors such as socio-economic background, gender, effort and behaviour may bias teachers' assessments (Bennett et al., 1993; Brookhart, 2011; Brown et al., 1998; Harlen, 2005a; Johnson, 2013; Klenowski & Wyatt-Smith, 2014; Mahuika et al., 2011). It is important that beginner teachers be made aware of these influences.

8.1.1.5 *Knowledge of quality in summative assessment*
Beginner teacher commentary indicated they had a tacit understanding of the principles of good assessment: validity, reliability, ethics and fairness, although they did not always use this terminology (see Section 5.5). When findings were compared to the eight components for making valid inferences (Crooks et al., 1996) there was evidence of each of the components in comments from most of the teachers. Nonetheless, the teachers said that at times they made ad hoc decisions in the administration, design and marking of tasks to match their own understandings of validity. As established in the literature, validity is a challenging aspect of summative assessment (Black et al., 2010; Lovett & Sinclair, 2005). Findings from this study add support to the call for ongoing professional development to redress weaknesses in this area (Black et al., 2010; Stobart, 2001).

Beginner teachers in this study were particularly aware of the need for reliability for NCEA assessment. As discussed in Section 8.1.1.4, the teachers found making judgements about student work challenging, and they did admit at times to allowing construct-irrelevant factors to influence their judgements, especially for non-NCEA assessment, decreasing the quality of assessment judgements. However, the moderation system in place for NCEA did provide them with a quality assurance safeguard.

In schools where beginner teachers did have agency in summative assessment, trade-offs between construct validity and reliability showed that they prioritised validity over reliability. Thus the first three requirements for dependability proposed by Harlen (2005a) (decisions about domain of knowledge, a valid sample of student work, and well-matched criteria) were prioritised by teachers working in junior science, and the final two requirements (procedures for reliable and unbiased application of criteria, and reporting and communicating) were less of a focus for the teachers. In general, the beginner teachers were less concerned with consistency and more concerned that junior students were given tasks that they could engage with meaningfully, in the hope it would stimulate ongoing interest in science.

The principles of fairness and ethical assessment featured strongly in the teachers' interviews and they all showed a tacit understanding of "do no harm" and "avoid score pollution", as suggested elsewhere as ethical guidelines (Green et al., 2007; Pope et al., 2009). They gave examples of practice they felt did not meet these requirements, mirroring the findings of other studies (Colnerud, 2006; Green et al., 2007; Tierney, 2014). Although they struggled to make decisions when they struck ethical dilemmas, there was evidence that they were cognisant of their need to increase their knowledge in ethics.

8.1.1.6 Summary

In summary, this section has discussed characteristics of summative assessment by focusing on beginner teachers' knowledge of purposes, what to assess, how to assess (strategies and task design), interpretation, and quality. This multifaceted knowledge base for summative assessment can be regarded as an amalgam. Moreover, it is drawn upon by beginner teachers during summative assessment decision making. Overall, beginner teachers were found to have a broad understanding of a wide range of summative assessment knowledge, although they were more confident in some areas than others.

However, a further characteristic of summative assessment is that decisions and action are not taken on the basis of knowledge alone. In the following section the personal and contextual factors that act as amplifiers and/or filters for what beginner teachers consider and choose to implement in practice are discussed.

8.1.2 Amplifiers and filters

As acknowledged in the Model of Teacher Professional Knowledge and Skill (Gess-Newsome, 2015), knowledge is personalised by teachers as it passes through their own lens of amplifiers and filters. This personalisation of teacher professional knowledge impacts teacher decision making and classroom practice. In this study four factors that acted as amplifiers and filters were identified: the sociocultural context (Section 8.1.2.1), teacher conceptions and background (Section 8.1.2.2), ethical matters (Section 8.1.2.3), and teacher emotional responses (Section 8.1.2.4). These amplifiers and filters were seen to mediate the translation of teacher professional knowledge to classroom summative assessment practice for different teachers in idiosyncratic ways.

8.1.2.1 Sociocultural context

This study found that the sociocultural context in which teachers learned and worked affected their learning and development of summative assessment literacy (see Section 6.3). The sociocultural context also acted as an amplifier and filter for teachers, affecting their assessment decision making (see Section 7.1.1). The context that had the greatest effect on the teachers was the school in which they were working. Teachers were able to identify aspects of school culture and practice that affected what they learnt and how they made summative assessment decisions e.g., school policies and practices (see Section 7.1.1). Beginner teacher commentary indicated their pedagogy and summative assessment decisions were affected by having to comply with their schools' assessment policies to the point of, at times, them enacting practices that were contrary to their own knowledge of the principles of quality assessment or that contradicted their own beliefs about assessment. These findings align with previous studies which have found that school policy does affect teachers' learning and behaviour, including decision making (Bullough, 1992; Peressini et al., 2004; Taber et al., 2011).

In this study school contexts were found to be one of two types. Firstly, some schools tightly controlled beginner teachers' summative assessment. In these schools teachers were expected to use the resources and time frames set by senior staff so they did not get the opportunity to develop their own assessment tasks, marking schemes, timeframes and so on. This meant the schools' priorities were acting as a filter for teachers. The vignettes of Bella and Ryan illustrate the lack of

experience gained by beginner teachers who are placed in these schools. A second type of school was less controlling, allowing teachers considerable freedom to assess students as they saw fit. Teachers working in these schools had much more agency in summative assessment decision making and were able to assess in ways consistent with their assessment beliefs and knowledge. Beginner teachers appreciated this and felt this freedom led to more development opportunities. The vignette of Susie illustrates a beginner teacher gaining experience through being placed in practicum schools like this.

As an example of how school culture can act as an amplifier/filter for beginner teachers, consider the case of Wiremu. Wiremu had his first practicum in a traditional and academically focused school which promoted academic excellence through formal assessment, and strictly controlled all aspects of summative assessment, giving beginner teachers no agency in the process. By the end of practicum in this school Wiremu seemed to have abandoned his earlier espoused beliefs which included prioritising oral assessment. Instead he talked about the need for conformity in summative assessment, treating students as scholars through the use of formal assessment (see Section 5.3.2). His second practicum was in a school context in which he had some autonomy to make summative assessment decisions. After time in this school he was found to return to his earlier beliefs. The shift in Wiremu's practice and espoused views illustrates the effects of contextual policy priorities and culture on a teacher's ongoing thinking and summative assessment decision making (see also Brown, 2011; Brown et al., 2011; Mahuika et al., 2011; McGinnis et al., 2004; Rogoff, 2008; Shulman & Shulman, 2004). As this example shows, teachers do react to different school cultures to the point where their developmental paths can be affected by school culture (Bullough 1992; McGinnis, 2004).

Given the bicultural context in New Zealand it is important that cultural elements of teacher summative assessment literacy are considered; however, there is currently very little research on this available in New Zealand. The studies by Mahuika et al. (2011) and Sun (2010) are exceptions. They support international studies that show that the cultural backgrounds of teachers affect their conceptions of assessment (Brown et al., 2011; Remesal, 2011; Seah, 2002). With respect to culturally responsive summative assessment, all the beginner teachers were aware

of its need in their multicultural classes, but none felt equipped to make well founded decisions. Unlike the findings of Mahuika et al. (2011), there was no evidence of beginner teachers regarding Māori and other minority students in deficit terms, instead they were all keen to learn how best to approach assessment with such groups.

8.1.2.2 *Teacher conceptions and background*

The conceptions and background of beginner teachers was discussed as a very influential lens that had a profound effect on their focus as they considered assessment and assessment decision making. Beginner teachers often spoke of their assessment actions in terms of their own beliefs and experiences (see Section 7.2). This finding echoes those from a wide range of studies which have reported that teachers' backgrounds, prior knowledge, beliefs, conceptions, theories and attitudes influence their learning and practice (Abell & Siegel, 2011; Buck & Trauth-Nare, 2009; Clark & Petersen, 1986; Ell et al., 2012; Flores & Day, 2006; Hollingsworth, 1989; Jones, 2010; Kennedy, 1999; Lortie, 1975; Stuart & Thurlow, 2000; Taber et al., 2011; Wang et al., 2010; Yilmaz-Tuzun, 2007). The pre-existing beliefs and backgrounds of teachers have been found to provide perspectives which "serve as culturally based filters to help make sense of the program content, their roles as student teachers, their observations of classrooms at work, and their translation of program content into teaching/ learning activities in the classroom" (Hollingsworth, 1989, p. 162). In this study three sets of beliefs in particular were found to be held by a number of teachers, and had an amplifying or filtering effect. Additionally, one teacher explained that he was profoundly affected by his own cultural beliefs, and these were seen to affect his assessment decisions.

Firstly, all teachers believed that summative assessment needs to be improvement-focused, and talked about their own actions in response to this belief, which resulted in them using summative assessment formatively. They implied this underpinned their motivation to learn about ways to provide useful feedback and to enhance their planning and contributed to actions such as re-teaching content, or planning to do things differently next year (see Sections 6.2.1.1, 8.1.1.1). They also talked of how they might improve the summative assessment experiences for students (see Section 6.2.1.2). The formative use of summative assessment has

262

been documented in a number of previous studies (Harlen, 2005b, 2006; McMillan, 2005; Timperley, 2009) and is a practice encouraged in the New Zealand setting. In this study teachers attributed these actions to their beliefs about making use of assessment to improve teaching and learning.

A second set of beliefs related to the importance of students' creativity, which beginner teachers felt was being ignored within science assessment. An effect of high stakes assessment is that the curriculum tends to be narrowed so that students are prepared for the assessment (Black, 2013b; McMillan, 2013). This certainly is a criticism that has been made of the NCEA system. Unless achievement standards specifically require assessment of some form of creativity, there is little evidence of it being assessed, as teachers feel compelled to focus on the criteria being assessed and ignore other aspects of the curriculum (Hume & Coll, 2008; Lock, 2005; Rawlins et al., 2005). Beginner teachers in this study talked about their commitment to encourage students to be creative. This demonstrates the amplifying feature of this belief, in that they were committed to acknowledge and encourage creativity even though this was not an expectation in most schools. The narrowing of the curriculum that occurs when teachers only focus on criteria from standards is linked to the third set of beliefs about student engagement, also seen to operate as amplifiers and filters.

The third set of beliefs seen to strongly affect summative assessment decision making in this study concerned the importance of student engagement in science. These beliefs were demonstrated to affect what the beginner teachers prioritised in their approaches to summative assessment. All beginner teachers wanted students to be interested and motivated to continue to learn science and felt that providing positive assessment experiences would encourage students to do so (see Section 7.2.2). This belief was linked to their concern about students' affective responses to assessment, and on occasion beginner teachers adjusted assessments to increase the positive outcomes for students.

As an example of the filtering effect of his belief about the importance of student engagement, Ryan voiced strong opposition to formal examinations from the beginning of the study and he prioritised learning about alternative assessment through internal assessments. He attributed his actions to his beliefs about the

potential demotivation, lack of engagement and damage for students who faced formal examinations. He expressed a lack of commitment to NCEA external examinations. His beliefs were strong and held with some passion, and did not change throughout the study (see Section 6.2.2.3). His case illustrated that beliefs can persist even after compelling new learning is provided through ITE, parallel to findings in other studies (Bransford et al., 2000; Dixon & Haigh, 2009; Edwards, 2003), and that these beliefs acted as a filter for his learning about formal summative assessment. When Ryan was in schools where he was forced to use formal examinations this caused him considerable stress.

Another aspect of teacher belief seen to act as an amplifier was cultural beliefs (see Section 7.1.2). For Wiremu oracy was an important element of his Māori cultural heritage. Wiremu trialed an oral summative assessment in English and te reo Māori while on practicum in his second school. He found it more challenging than expected, but said he would continue to trial oral assessment strategies, given its cultural significance for Māori (Edwards & Edwards, 2016). This finding aligns with those of other researchers who found the formation of teacher identity is a complex and culturally based process, which impacts on decision making (Chong, Ling, & Chuan, 2011; Danielewicz, 2001). In this case the amplifier of teachers' cultural background was seen in action as Wiremu focused on an element of summative assessment that is prominent in his culture.

The amplifying or filtering effect of beginner teachers' conceptions and backgrounds has featured in other studies focused on early career teaching: in ITE programmes (Kagan, 1992), throughout ITE programmes (Postlethwaite & Haggarty, 2012), and when transitioning to teaching (Haggarty & Postlethwaite, 2012). Although Levin and He (2008) found, in their study of preservice teachers, that an ITE programme can significantly influence teacher beliefs at least until the end of the ITE programme, they also conceded that this does not mean that other things do not contribute to teachers' beliefs and actions, and serve to act as filters for what they might learn through ITE.

8.1.2.3 *Ethical matters*
Ethical matters relate to an important subset of teacher beliefs, and beginner teachers in this study inferred they were particularly salient as amplifiers and

filters. Specifically, findings indicate that all teachers felt a strong obligation of ethical care towards their students, treating them with respect and care. They also highlighted their beliefs regarding the need for honesty and fairness in assessment. These beliefs led them to make decisions that affected their summative assessment practice.

A clear finding which illustrated the amplifying and filtering effect of beginner teachers' ethical beliefs concerned their beliefs about their obligation to treat students with respect and care (Section 7.3.1). This obligation meant they prioritised the care of students and voiced a commitment to be inclusive of all learners. The ethical principle of "do no harm" was very evident in all teachers' interview comments and is seen as a core ethical principle in a number of professions (Green et al., 2007). The beginner teachers' beliefs regarding treating students with respect and care led a number of them to make decisions to provide more flexible summative assessment design, irrespective of whether this was sanctioned by their schools.

A second finding relating to teachers' views of what ethics are was the need for honesty. The findings indicated that beginner teachers voiced their need to be honest in their dealings with students, truthfully appraising the evidence of student learning as they made assessment judgements (see Section 7.3.2). They raised issues that troubled them such as their observations of teachers teaching to the test or not following moderation protocols. Beginner teachers in this study did not admit to writing positively biased report comments, but given the difficulty they found in being honest about students' achievement because of their emotional engagement with the students (see Section 7.4) it could be anticipated that they would find it challenging to be completely honest, as has been found for teachers in other studies (Hattie & Peddie, 2003; Robinson & Timperley, 2000).

All beginner teachers emphasised their beliefs in fairness as an ethical issue in summative assessment. As found in other studies, beginner teachers had various interpretations of fairness that centred around ideas relating to equality or equity (Stobart, 2005; Tierney, 2013, 2014). As the study progressed, these became more focused on issues of equity (see Section 6.1.3.2). Beginner teachers raised issues of fairness for students who found it difficult to show what they knew in an

assessment (e.g., for students with low literacy levels, ADHD, English as an additional language). They talked about how they made decisions in an attempt to ameliorate the impact of summative assessment tasks in order to make assessment more equitable for their diverse students (see Section 5.3.2). This is in line with calls for teachers and schools to move towards supporting diverse students' learning rather than treating students equally (Lyon, 2013c; Mahuika et al., 2011; Tierney, 2013).

Pope et al. (2009) found that experienced teachers faced ethical dilemmas often concerned with grading, standardised testing and special populations, and that conflicts often arose when there were differences between what teachers saw as the needs of their students and institutional demands. In their study, as in this one, it was found that teachers applied ethical principles based on their own beliefs in ways that they saw fit, and they found that their perception of the demands of schools sometimes made their dilemmas more difficult to deal with (Pope et al., 2009). The prevalence of institutional demands causing ethical conflict is consistent with the findings of a number of international studies where the strong influence of policy has been seen to constrain classroom assessment practices and make teachers feel compromised by institutional requirements (Black & Wiliam, 2005; Colnerud, 1997).

International studies have found teachers to have limited understanding of principles of ethical decision making and not be equipped to deal with ethical decisions and dilemmas (e.g., Green et al., 2007; Plake & Impara, 1997; Pope et al., 2009; Tierney, 2013). However, in this study the beginner teachers did explain how they applied their own ethical beliefs when making decisions about how to summatively assess students. There have been calls to focus more explicitly on ethical decision making in ITE courses (Gipps, 1994; Popham, 2000; Shapira-Lishchinsky, 2011; Strike, 1990). It would be important that this was done in a way that acknowledged the amplifying and filtering effects of beliefs already held by beginner teachers.

Teacher emotions is closely linked to ethical matters as amplifiers and filters in assessment decision making. This is discussed next.

8.1.2.4 *Teacher emotional responses*

Teacher emotion acted as an amplifier or filter for beginner teachers' assessment decision making. All teachers communicated a sense of emotional engagement with their students with respect to assessment. They also discussed the emotions associated with being judged as teachers.

Beginner teachers experienced a range of emotions as they assessed students (see Section 7.4.1). A number recalled their own feelings of anxiousness when being assessed and projected these feelings onto their students. The amplifying and filtering effects of these emotions were evident in teachers' assessment decisions. For example, those teachers who remembered feeling anxious and worried when facing examinations tried to avoid or reduce their use of examinations, or at least tried to mitigate stress for students (see Section 6.2.2.3). Research evidence does point to students reacting emotionally to being assessed and receiving feedback (Vogl & Pekrun, 2016), so teachers are right to try to reduce negative emotions in their students. To this end, beginner teachers redesigned tasks to try to make them more enjoyable and adjusted conditions to enable students to do better in assessments as mentioned in Section 8.1.2.3.

Seven of the beginner teachers experienced feelings of fear or worry during the process of marking or grading student work, particularly when they considered the effects the grades might have on the students. For example, in apparent contradiction to their beliefs about honesty, they talked about their own decisions to sometimes relax criteria for particular students when marking, adjusting timing, and giving hints to some students in summative assessment events, particularly in non-NCEA assessments (see Section 7.4.1). These behaviours were as a response to their concern about students' emotional responses to assessment results and could be interpreted as dishonest, showing that the beginner teachers' concerns for students' emotions were prioritised over their beliefs relating to honesty. These sorts of actions have also been reported in other research over a number of years (Brookhart, 1999; Brookhart, 2013b; McMillan, 2003; Randall & Engelhard, 2010), and can be attributed to the pressures that stem from high stakes assessment in any system focused on accountability as well as teachers' own motivations and beliefs (Amrein-Beardsley, Berliner, & Rideau, 2010; Nichols & Harris, 2016).

This study illustrates the link between emotion and cognition, and the findings are in line with other studies which highlight the emotional work of teachers and the strength these emotions bring to bear on teachers' decisions (Näring et al., 2006; Nias, 1989; Soini, Pyhältö, & Pietarinen, 2010). Kelchtermans (2005) argues that "emotion and cognition, self and context, ethical judgement and purposeful action: they are all intertwined in the complex reality of teaching" (p. 996) and this complexity is evident in the dilemmas faced by beginner teachers.

Being emotionally engaged with students and their progress can cause teachers to become disappointed or frustrated (O'Connor, 2008), as was evident in this study. For example, all beginner teachers described times when they felt disappointed or angry but powerless as they observed their students underperforming or opting out of summative assessments (see Section 5.1.5). This behaviour has been found to be a feature for some students in NCEA assessment as they focus on achieving the minimum credits required (Graham et al., 2010; Hipkins et al., 2005; Meyer et al., 2009). Shapiro (2010) argues such feelings are a result of the tension between teachers' concerns of an intellectual nature and their emotional responses. This sort of tension adds to teachers' emotional labour (Isenbarger & Zembylas, 2006) and may well be more keenly felt in beginner teachers given their lack of experience.

Being judged as a teacher also generated emotion for beginner teachers, as they felt that their performance as teachers was judged based on their students' grades (see Section 6.4.2). This produced positive emotions when their students did well and negative emotions when students did poorly in assessments, and at times generated a level of self-doubt for the teachers. Teachers felt vulnerable when their students' grades were scrutinised by others. Vulnerability has been described as "feeling that one's professional identity and moral integrity, as part of being 'a proper teacher', are questioned" (Kelchtermans, 2005, p. 997) and can arise in teachers when they do not understand the limits of their professional efficacy. It is important for beginner teachers to have realistic expectations of their influence.

8.1.2.5 Summary
In summary, this section discussed the findings with respect to four amplifiers and filters in operation for teacher learning and decision making about summative

assessment: sociocultural context, teacher conceptions and background, ethical matters, and teacher emotional responses. Teacher commentary indicated that these amplifiers and filters affect their assessment learning and decision making because they provide a lens through which beginner teachers see and experience their work. The presence of amplifiers and filters provides one explanation for why individual teachers' development and use of knowledge bases is not altogether uniform. Apparent conflicts in teacher behaviour can be observed and explained as teachers make decisions influenced by their own mix of amplifiers and filters in action.

Overall, teacher summative assessment literacy can be seen as an amalgam of knowledges that informs and shapes teachers' assessment decision making. Furthermore, this amalgam of knowledges is unique to each teacher. To add complexity, personal and sociocultural factors act as amplifiers and filters for what teachers learn and choose to implement in practice. This means that teacher assessment decisions are based on more than their knowledge about assessment; an understanding of teachers' knowledge alone is not enough to understand their actions. As has been found in this study, teachers' practice therefore may be idiosyncratic for each teacher depending on their own beliefs, interpretations and context.

8.2 The development of summative assessment literacy over time

This section responds to the second research question, "In what ways does beginner secondary science teachers' summative assessment literacy develop during their formal teacher education and beginner teaching experience?" Patterns of summative assessment literacy development were evident through the use of the SALRubric (see Section 4.5.2), by which teachers' summative assessment literacy development across ten dimensions was tracked over time (see Section 6.1–6.2, Figure 4, and Appendix H).

For all beginner teachers in the study, considerable shift in their development was observed over time as they developed deeper knowledge and understanding of the practice of assessment. This sort of development is consistent with the findings of many other studies which have documented teachers' development in assessment literacy through ITE and early teaching experiences (DeLuca & Klinger, 2010;

269

Edwards & Cooper, 2012; Eyers, 2014; Graham, 2005; Lyon, 2013b; Smith & Lev-Ari, 2005; Smith et al., 2014; Volante & Fazio, 2007). There was evidence of a shift in all of the dimensions identified in the SALRubric, though the development was not uniform across teachers or time. For each beginner teacher in this study an individual and distinct pattern of development was observed through the use of the SALRubric (see Figure 4 and Appendix H). Similar patterns of development across all teachers were evident for some knowledge dimensions, but not all. For example, the development of *ability to describe assessment* and *knowledge of assessment purposes* was similar across teachers, with most making changes at similar times. However, the patterns for *knowledge about what to assess* and *knowledge of assessment strategies and design* were quite varied, with teachers' trajectories being very different from each other. The variability in developing a knowledge base is unsurprising given the large number of factors which affect teacher development.

Beginner teachers' development with respect to *NCEA assessment* took time, and in general they lacked confidence about their role in NCEA assessment. The technical aspects of using the NCEA system proved most challenging for teachers. The NCEA system has been found to be demanding for teachers (Hipkins, 2010, 2013). In line with the findings of Smith et al. (2013) and Anthony and Kane (2008) the teachers in this study lacked confidence in the use of NCEA through their ITE year and needed practice time in schools to gain confidence and expertise. However, their confidence increased gradually over time, with teachers reporting a significant gain in confidence in NCEA assessment after their first six months of teaching. Teachers' concern about teaching to the test did seem to play a role in the delay of their development of an understanding of *preparing students for standards-based assessment* (see Section 6.1.2.2). The teachers' own observations as well as research evidence (Hume & Coll, 2010; McMillan, 2013; Moeed & Hall, 2011) indicates that teaching to the test can be an issue within the NCEA system, particularly for internally assessed standards given the teacher control of the whole process.

Teachers appeared motivated to *use summative assessment formatively* from the start of the study, and the SALRubric judgements indicated their early uptake of the formative use of summative assessment (see Section 6.1.2.3). This is in

270

contrast to other New Zealand studies focusing on primary teacher assessment development, in which teachers took time to develop an understanding of how to use summative assessment formatively (Eyers, 2014; Hill et al., 2013; Smith et al., 2014).

Beginner teachers' summative assessment literacy developed at variable rates in both dimensions of *understanding assessment consequences*, and *fairness*. These were areas in which teachers were more emotionally involved (Section 8.1.2.4). The public nature of teachers' work, including assessment, means that decision making can be complex to deal with (Fried et al., 2015; Näring et al., 2006; Zembylas, 2004). Over the course of the study teachers developed a deeper understanding of the nuanced nature of the consequences of summative assessment and of fairness. They developed their ability to make decisions which would influence these dimensions in a positive way.

Idiosyncratic patterns of development or trajectories for teachers have been noted in previous studies (Aydin et al., 2015; Bullough, 1992; Hutner & Markman, 2016; Lyon, 2013b; McMillan & Nash, 2000; Park & Chen, 2012; Park & Oliver, 2008; Remesal, 2011). As suggested in the literature, teachers' beliefs, orientations, prior knowledge, views of learning and context act as amplifiers and filters in teachers' learning, affecting what they notice and how they learn (Campbell, 2008; Eraut, 1994, 2009; Gess-Newsome, 2015). These affect the way individual teachers accept or reject ideas and suggestions, as well as how they make classroom decisions. So although the teachers in this study were exposed to the same ITE programme, their summative assessment literacy developed in different ways. For example, in this study Ryan was a teacher who had particularly negative views about formal summative assessment and its impact on students. As can be seen in the radar diagram of his development (see Figure 4), Ryan's development took longer than others, and shift was not seen in his competency for a number of dimensions for a longer time. It is likely that his personal beliefs and his own memories of emotional experiences linked to being assessed affected his openness to new ideas and his willingness to learn about summative assessment. Additionally his experience in practicum schools where he was given little opportunity to be involved in summative assessment did little to help his development.

Of particular note was the general pattern of change across teachers, which indicated that practicum appeared to facilitate or be a catalyst for summative assessment literacy development in most dimensions for most teachers. The main shifts in competence were observed between Interview 1 and 2, and Interview 3 and 4 which coincided with practicum (see Table 2). Practicum is known to be very important for teacher learning as it has been found to provide teachers with opportunities to trial ideas, develop knowledge and skill and gain confidence (Bullough, 1992; Haigh et al., 2008; Smith & Jang, 2011; Smith & Lev-Ari, 2005; Winterbottom, Brindley, et al., 2008). Beginner teachers in this study appreciated the opportunity to put what they were learning into practice (see Section 7.4.3) as well as the learning opportunities gained through discussions with mentor teachers, and showed considerable development as a result. As found in other studies (e.g. Bullough, 1992; Grudnoff, 2007, McGinnis et al., 2004), teachers' school experiences did vary considerably and the quality of AT supervision and level of support from the schools did impact on the learning experience for beginner teachers. It was noted that teachers who worked in schools which gave them very little autonomy had a slower developmental trajectory for *knowledge of assessment strategies* and *knowledge of what to assess* (see Section 6.1.1.3, 6.1.1.4). It has been shown that when teachers make links between what they learn in university courses and their practicum experiences, this heightens the benefit of both (Adoniou, 2013; Allen, 2009; Bullough, 1992; Cheng et al., 2012; Darling-Hammond, 2006b). In this study some teachers were aware of making these links and talked specifically about doing so (see Section 7.4.3). Susie's vignette shows that serendipitous practicum placements can provide a beginner teacher with multiple opportunities for development in summative assessment literacy.

Unique patterns of development are often explained by considering teachers' personal sets of beliefs and teaching experiences. For example, in Lyon's (2013c) study of the development of equitable assessment expertise in science teachers it was found that experiences such as teachers' interactions with students and the topics they were able to teach on practicum fed into their learning. Lyon's (2013c) findings emphasised the complexity in developing assessment expertise and the idiosyncratic trajectories of development that resulted for teachers over time. This

study adds a new perspective to Lyon's work by focusing on summative assessment over a wider range of dimensions (ten dimensions rather than three) and in a very different context in which teachers have more agency, working in a standards-based assessment system. It elaborates on the complexity of teacher development and extends the ideas of learning progressions of assessment expertise as proposed by Lyon (2013a).

The findings also affirm the work of Schneider and Plasman (2011), who found that science teachers moved from using a very limited set of assessment strategies to using a wider range of assessment knowledge and also developed in their understanding of the use of assessment. As might be expected, by the end of this study none of the eight teachers had reached Expert; most were judged as Competent for most dimensions (See Appendix H). Therefore teachers could be described as demonstrating emerging adaptive expertise and preparedness to assess by the end of their ITE, in line with the findings of other researchers (e.g., DeLuca & Klinger, 2010; Eyers, 2014; Smith et al., 2014).

8.3 Contributors to development of summative assessment literacy over time

This section responds to the third research question, "What do beginner secondary science teachers identify as contributors to the development of their summative assessment literacy?" In this study three specific elements were identified and these have been discussed in the following sections: people (Section 8.3.1), learning activities (Section 8.3.2), and teaching experiences (Section 8.3.3).

8.3.1 People
Three groups of people had a significant impact on beginner teachers' development of summative assessment literacy: ITE lecturers, other classroom teachers, and ITE classmates (see Section 6.3.1).

Firstly, university lecturers from the ITE programme were identified as contributors to beginner teachers' summative assessment literacy as well as to their general confidence and willingness to take risks when teaching. However, from a large group of possible lecturers it was clear that three lecturers were particularly influential, as these same three lecturers were identified by every beginner teacher in the study. The beginner teachers valued the relationships that

these three university lecturers built with them and they talked of the positive ways the lecturers related to them showing a genuine interest in them. Based on this relationship, the content that these lecturers taught and the way they modelled good pedagogy was well received and appeared to create real and memorable impact.

The importance of the teacher-learner relationship has been acknowledged in all sectors of education including teacher education (Giles, 2010; Korthagen, 2010; Loughran, 2013; Shoffner, 2009). In this study genuine personal encounters between teacher educators and their students were seen to open up learning spaces and encourage engagement with what was being taught. Giles (2010) talked of the need for teacher educators to be attuned to the "play" of the interactions and relationships they have with student teachers showing phronesis in the way they relate, and Shoffner (2009) notes that the affective domain is also important in teacher preparation. From the rich descriptions offered by the beginner teachers in this study, and my own observations in their university classes, it was evident that these three lecturers were particularly skilled in doing this.

Secondly, ATs and mentor teachers were reported as playing a very important role in the development of teachers' summative assessment literacy. On practicum beginner teachers highlighted the advice from ATs which they held in higher regard than advice and feedback from the visiting university lecturers (see Section 6.3.1). Flores (2003) emphasises the way teachers draw on past experiences and the knowledge of others as they construct their teacher knowledge, and how this knowledge is developed socially by individuals with others. There were occasions, however, when the ATs were unhelpful or confusing for the preservice teachers, particularly over NCEA summative assessment. These findings are similar to the findings of other studies which emphasise the variability of practicum experiences beginner teachers get (Bullough, 1992; Eyers, 2014; Haigh et al., 2008; Volante & Fazio, 2007). Relationships with ATs can enable or hinder teachers' development when on practicum, with good communication as a result of these relationships being essential for positive outcomes (Haigh et al., 2008).

Mentor teachers were mostly found to be helpful and supportive by the beginner teachers in their first schools. However, it was clear from the teachers' comments

274

that the levels of trust involved in their relationships with mentors dictated their openness and the sense of support they received. Some mentors closely supervised beginner teachers and gave them little freedom with respect to summative assessment decisions. This type of mentoring relationship was found to be less helpful as the teachers preferred to be able to manage summative assessment for their own classes. A number of researchers have highlighted the role of mentor teachers in beginner teacher development and identified that on-site experienced mentors play a pivotal role in beginner teacher development (Aitken & Mildon, 1992; Bullough, 1992; Volante & Fazio, 2007). In her work with New Zealand teachers, Grudnoff (2007) also found that the professional and social interactions with mentors were vital in teachers' transition as they began teaching. These interactions provide emotional as well as practical and technical help to beginner teachers. This study suggests that the summative assessment literacy of beginner teachers was best served by mentor teachers who were supportive and available, knowledgeable about summative assessment and willing to let the teachers take ownership of their assessment decisions.

Thirdly, teachers identified their ITE classmates as contributors to their summative assessment development (see Section 6.3.1). During university class activities, beginner teachers found it very fruitful to work collaboratively with their classmates, discussing their knowledge and understanding of summative assessment. They also appreciated learning about each other's practicum experiences and sharing concerns and worries in a non-judgemental space, usually in their science education course. This sort of discussion and collaborative learning is an important element in beginner teacher development (Loughran, 2013), with perspective-building conversations being identified as one of four specific pedagogical constructs that contribute to learning about assessment (DeLuca, Chavez, Bellara et al., 2013). Collaboration requires teachers to articulate and synthesise their perspectives based on their own knowledge, experience and understanding of theory. As seen in this study, such conversations allow teachers to gain multiple perspectives and scaffold each other's learning as they discuss readings, classroom scenarios regarding assessment and work on classroom activities together.

8.3.2 Learning activities

In this study there were a small number of course-based learning activities that had a profound impact on beginner teachers (see Section 6.3.2). The teachers all independently identified the same learning activities as contributors to their summative assessment knowledge and confidence. These activities were structured workshops in a science education course which focused on linking unit planning, lesson planning, curriculum and assessment; short modules on developing marking rubrics; workshops focusing on literacy across the curriculum collaboratively; and marking real student work against NCEA standards.

The beginner teachers in this study had no specific assessment course with their ITE programme, so teaching about summative assessment was embedded into other courses. Assessment courses have been found to be effective in helping teachers develop knowledge and understanding of assessment (DeLuca et al., 2010; Eyers, 2014). Additionally, course embedded projects have also been found to be useful in teacher development (Hume, 2006; Hume & Berry, 2010; Youngs & Bird, 2010). The literature is clear that it is the cohesion of the ITE programme linking university class work and teaching practice together that is critical for effective and timely teacher development in that it enables teachers to "reinforce, apply and synthesize concepts they are learning" (Darling-Hammond, Hammerness, Grossman, Rust, & Shulman, 2005, p. 401). In the modules or sessions identified by beginner teachers the links to summative assessment practice were made explicit.

All identified learning activities were explicit in linking theory and practice and all included an element of authenticity. A series of activities which involved the teachers developing a rubric that matched an NCEA achievement standard and then marking real student work with their rubric was highlighted. Whether in a stand-alone assessment course or an integrated programme, teachers need experience in designing, scoring and interpreting assessment tasks as an important part of their ITE (Shepard et al., 2012), and the authentic use of the rubric in collaborative marking of student work was powerful. Similarly, an activity focused on unit planning for real teaching allowed lecturers and students to make the links to classroom practice. Workshops about literacy across the curriculum had similar impact as the lecturer carefully linked theory to practice. These

276

activities were authentic and could be applied directly to the teachers' own teaching practice. The use of authentic tasks in university learning activities or for assessment within university courses emphasises the connection between assessment theory to teaching practice (DeLuca, Chavez, Bellara et al., 2013).

8.3.3 Teaching experiences

Teaching experiences, whether on practicum or in their first schools, were identified as a very important contributor to the beginner teachers' development of summative assessment literacy (see Section 6.3.3). The beginner teachers appreciated being involved in summative assessment when working in practicum schools. As discussed in earlier sections (Section 8.1.2.3 and 8.1.2.4), this work was at times challenging. However, teachers valued the opportunities they were given and talked about how their practicum helped them to consolidate what they were learning at university, and evidence of this was clear through use of the SALRubric (see Section 6.1).

Echoing other studies, practicum is seen as a critical element within ITE that can lead to a better developed confident and knowledgeable teacher or a less confident teacher, depending on the experience (Smith & Jang, 2011; Smith & Lev-Ari, 2005; Taber et al., 2011; Winterbottom, Taber et al., 2008). As has been found previously, teachers in this study felt more able to participate in the full role of teacher, including as an assessor, through the knowledge and confidence they gained from teaching experiences. They appreciated the learning they were able to do through their practical application of theoretical knowledge and through reflecting on the experiences to see links to theory, in much the same way as described by Tsui (2009).

With respect to summative assessment, development teachers were particularly appreciative of the exposure they received regarding NCEA and the opportunities they were given to mark work or oversee NCEA assessments while on practicum and teaching in their schools. They viewed this as crucial to their preparation and work as secondary school teachers in New Zealand. By getting to know the content and how to assess a range of achievement standards they felt they were building up a bank of knowledge about the standards and about standards-based assessment. Authentic opportunities such as these provide teachers with the

opportunities to integrate their theoretical knowledge of assessment with classroom practice in order to further their development (Graham, 2005; Nolen, Horn et al., 2011).

8.4 Summary

This study has contributed insights into the characteristics of summative assessment literacy, its development and what beginner teachers identify as contributors to its development. Summative assessment literacy was found to be an amalgam of knowledges that informed beginner teachers' summative assessment decisions. The knowledges were wide ranging and included knowledge of assessment purposes, what to assess, summative strategies and task design, assessment interpretation and the principles of quality assessment. Teachers identified a range of factors that affected their summative assessment decision making, namely amplifiers and filters that acted as a lens for the teachers. The amplifiers and filters that were identified in this study were both personal and contextual factors: sociocultural context, teacher conceptions and background, ethical matters and teachers' emotional responses. They affected the way teachers approached summative assessment and at times they caused a change in teacher response because of the way decision making priorities were affected by their personal views of summative assessment. Sometimes the responses or decisions that beginner teachers made appeared contradictory to their espoused beliefs because of the effects of context, for example.

Teachers learnt about summative assessment and developed summative assessment literacy over time throughout their ITE year and as they commenced in their first school of employment. As shown through the use of the SALRubric, their summative assessment literacy developed in idiosyncratic ways. The development of teacher summative assessment literacy appeared to occur over time, with the input of learning during practicum noted as a very important catalyst to its development. Beginner teachers identified people, learning activities and teaching experiences that contributed to their summative assessment literacy development.

The next chapter draws conclusions and suggests implications, limitations and possible future research based on the study.

278

Chapter 9: Conclusion

Chapter 9 comprises a summary of the key points of the study. The major implications for policy makers, ITE lecturers, schools and new teachers are discussed. An outline of the limitations of the study and suggestions for further research are presented.

The aim of this research was to investigate summative assessment literacy and its development in beginner secondary science teachers in New Zealand. As the literature reviewed in Chapters 2 and 3 suggests, assessment literacy is a necessary component for all teachers, but its development is not straight forward. In the New Zealand secondary school context, the demands on teachers because of their involvement in high stakes assessment means that their summative assessment literacy is of particular importance. However, the literature reveals that few studies have focused specifically on assessment literacy development over longer time periods. This gap in the literature, along with my personal experiences of working with secondary school teachers in New Zealand, led to this research topic.

A qualitative study was considered the most suitable research design in order to give a detailed understanding of the characteristics and development of summative assessment literacy over time. The summative assessment literacy of eight beginner science teachers was tracked from the start of their ITE programme until six months after they had commenced their first teaching positions. Data were gathered over this approximately 18 month period via a questionnaire (used three times during the study), a series of five semi-structured interviews for each teacher, observations in university course classes and the collection of artefacts deemed by the teachers as representing their summative assessment practice.

Content analysis of the data gathered from the questionnaires, interviews, observations and artefacts was used to identify trends and patterns. Models of science teacher assessment literacy (Abell & Siegel, 2011) and of teacher professional knowledge and skill (Gess-Newsome, 2015) were used as an initial framework for this process. Drawing on literature and New Zealand's distinctive policy and practice as well as initial findings, a Summative Assessment Literacy

Rubric (SALRubric) was developed to track the beginner teachers' summative assessment literacy development over time. Use of the SALRubric allowed development of dimensions of summative assessment literacy to be documented and visually represented in radar charts. This detailed analysis provided critical insights into the possible nature and development of summative assessment literacy in beginner secondary science teachers.

The following research questions guided the study:
a) What are the characteristics of summative assessment literacy as perceived by beginner secondary science teachers?

b) In what ways do beginner secondary science teachers' summative assessment literacy develop during their formal teacher education and beginner teaching experience?

c) What do beginner secondary science teachers identify as contributors to the development of their summative assessment literacy?

The next section summarises the answers to these questions.

9.1   Summary of the study

Addressing research question one, this study provided insight into two major characteristics of summative assessment literacy. Firstly, summative assessment literacy was found to be an amalgam of knowledges used to inform teachers' summative assessment decision making. Teachers demonstrated knowledge of the various purposes of summative assessment, of what to assess, of summative assessment strategies and task design, summative assessment evidence interpretation, and the principles and conditions needed for quality summative assessment. The amalgam was drawn upon, as teachers considered what strategies were appropriate to apply in a given situation and provided a basis for summative assessment decision making.

Secondly, both contextual and personal factors acted as amplifiers and/or filters for what and how teachers chose to assess in their classrooms. Contextual factors included external influences such as school policy and practice, and national policy. School culture was another important contextual factor that influenced the

teachers. Personal factors included teachers' background, conceptions and beliefs as well as their emotional responses. In particular, beliefs about assessment being improvement focused and the need for positive engagement in science led teachers to act in certain ways. Beliefs about the ethical treatment of students, and fairness, were also seen to affect teachers' summative assessment decisions about what to prioritise. The emotions beginner teachers experienced also acted as amplifiers and filters. While many of these amplifiers and filters were found in common for the beginner teachers in the study, their effects were distinctive and idiosyncratic for each teacher, reflecting the complexity of the development process.

It was apparent from this study, therefore, that beginner teachers' summative assessment literacy requires consideration of both their summative assessment knowledge as well as elements of their personal lenses which act as amplifiers and filters, causing teachers to approach the learning and application of new knowledge in their own ways. Overall findings confirm that the use of summative assessment knowledge is not as simple as applying knowledge or learnt strategies, but is bound together with the personal lens of the teacher. Teacher commentary indicates decisions are made based on the prioritising of different elements and are context specific.

The second research question addresses development over time. This study provided evidence that beginner teachers' summative assessment literacy develops in an idiosyncratic manner and that all show significant development over the 18 month period. It was possible to track beginner teacher development over time and across dimensions of summative assessment. For each teacher an individual and distinct pattern of development was observed, and the rate of change was quite variable for teachers and dimensions. In general, teachers moved from a novice understanding to the stage where they could summatively assess in their first school with some confidence, although they still required support, particularly with respect to aspects of NCEA and summative assessment task design. The most significant shifts of summative assessment literacy for all teachers was observed over practicum experiences, indicating that these had particular influence on their learning.

Addressing the third research question, the beginner teachers articulated a range of relationships, activities and experiences that contributed to their development. These teachers valued the relationships they built with a subset of the ITE lecturers and university learning activities that focused very specifically on summative assessment. They felt their teaching experiences in schools with ATs and mentor teachers was invaluable for their summative assessment literacy development.

9.2 Implications of this study

In New Zealand the goal of having assessment capable teachers, students and other stakeholders in the education system in order to promote system-wide improvement within and across all layers of the schooling system has been affirmed by the New Zealand Ministry of Education (2011). Based on the findings of this study, there are a number of implications for policy makers, ITE lecturers, schools and beginner teachers with respect to the development of teacher summative assessment literacy.

9.2.1 Implications for policy makers and government agencies

Educational assessment policy makers in New Zealand set the direction and requirements for teachers who are responsible for assessing students. There has been a prioritisation of formative assessment in education policy over the last 20–30 years, with less focus on summative assessment. Findings in this study illustrate the complexity involved in summative assessment literacy and its development for teachers. It is incumbent upon policy makers to acknowledge the knowledge required for quality summative assessment and the challenges that teachers face when making summative assessment decisions. Given that assessment literacy is a core professional requirement for teachers, it is important that educational assessment policy reflects the requirements of summative assessment including knowledge of the principles of quality assessment design and use (validity, reliability, ethics, fairness) alongside the importance of its formative use. Likewise, it is essential that government agencies ensure that ITE programmes have a component that focuses specifically on quality summative assessment and the development of summative assessment literacy in teachers. In particular, for secondary ITE programmes the requirement for NCEA specific assessment knowledge and experience could be emphasised. Additionally, the

ongoing provision of government funding for professional learning and development in this topic would ensure more clarity and a shared understanding of quality assessment for teachers

9.2.2 Implications for ITE providers and lecturers

A number of direct implications for ITE programmes and ITE lecturers were highlighted in this study.

Firstly, there were a number of knowledge areas which seemed to be lacking for beginner teachers. It is recommended that ITE providers provide programmes that allow beginner teachers to develop a robust knowledge base for summative assessment. As well as areas of strength in the ITE programme, this study showed knowledge areas which had little coverage e.g., data literacy for aggregated data, task design, choosing the most appropriate content to assess, moderation, student motivation and so on. A thorough review of an ITE programme should allow such areas lacking to be identified and directly addressed.

A further key implication is that it is incumbent upon ITE lecturers to develop an understanding of summative assessment literacy development that extends beyond the knowledge that teachers need to draw on, to include the impacts of amplifiers and filters on beginner teachers' learning about summative assessment and associated decision making. Beginner teachers can be made aware of the effect of amplifiers and filters as well as be given opportunities for personal reflection on the sets of amplifiers and filters that may affect and inform their own assessment decision making. By making the process of assessment decision making explicit, lecturers will help beginner teachers better negotiate what they need. Additionally, assessment decision making was found to affect teachers emotionally, and was especially challenging when teachers dealt with ethical dilemmas. ITE lecturers are in a position to alert teachers to the sorts of dilemmas they might encounter and the likelihood of potential emotional experiences. Emotional and pastoral support for beginner teachers when they face these challenges as part of their professional work would also be beneficial.

It is important that beginner teachers get assurances that as they progress through ITE and start teaching in their first schools they can expect their summative assessment literacy to develop over time. As found in this study, practicum is

likely to be important to their development of summative assessment literacy. The SALRubric was found to be a useful tool for tracking the development of summative assessment literacy over ten dimensions. There is scope for the SALRubric to be used by beginner teachers as a tool for self-assessment. Through its use beginner teachers could better understand and monitor their own development and make decisions about personal learning foci. Similarly, lecturers could make use of or adapt the SALRubric as a formative assessment tool. Some exemplars would help the users of the SALRubric to more clearly understand the boundaries in the rubric categories. As the SALRubric was developed in a New Zealand context, some of the dimensions may not be so useful in international contexts, but it could be redesigned or adjusted to better fit the context in which it is being used.

This study showed the importance that beginner teachers placed on the relationships they developed with ITE lecturers. This being the case, ITE lecturers are best placed to prioritise the investment of their time to develop relationships with the preservice teachers, rather than just providing teaching content in an impersonal way. The study clearly showed that beginner teachers felt they learnt best from lecturers who they related to at a personal level. They felt these lecturers modelled good teaching practice and this made the teachers more receptive to activities from these lecturers. This emphasises the social nature of the teacher-learner relationship and counters calls for large ITE classes in which the relationships with lecturers are less likely to develop.

The study demonstrated the importance of practicum within ITE for preservice teachers' development and reinforces the need for quality practicum experiences which allow teachers to practice and experiment within summative assessment so that they can put into practice what they are learning, as discussed below.

9.2.3   Implications for schools

School contexts were found to have an important influence on the development of beginner teachers' summative assessment literacy, and, as such, it is important that schools assist beginner teachers by providing them with opportunities to make assessment decisions and experience a wide range of summative assessment practices, such as NCEA marking or moderation meetings. Beginner teachers

need some freedom to make summative assessment related decisions in a supported way, rather than having all aspects of summative assessment controlled by others. This will allow the teachers to learn what works best, thus developing their own PCK.

The selection of ATs and mentor teachers for beginner teachers is important and requires careful consideration as these people have been found to have a substantial impact on teachers' development of summative assessment literacy and on their confidence and self-efficacy. In particular, it is important that ATs and mentor teachers understand the range of dimensions of summative assessment literacy and the best ways to give high quality feedback to beginner teachers. The SALRubric, or a modified version of it, could be used in schools to help them understand the trajectories beginner teachers take and to provide focus for feedback on summative assessment practice. Positive mentoring which encourages risk taking is very important, as is giving teachers the confidence and opportunities to try out things for themselves.

It is necessary that those working in schools recognise and support beginner teachers facing ethical dilemmas, giving them safe opportunities to discuss these as they arise. Equally important is their understanding of the part emotion can play in beginner teachers' summative assessment practice. Making such elements explicit could help beginner teachers negotiate development in this area.

9.2.4   Implications for beginner teachers

Beginner secondary teachers are aware of their need for summative assessment literacy, particularly with respect to NCEA assessment. It is important that they purposefully develop summative assessment knowledge as well as an awareness of the range of amplifiers and filters that affect their decision making and their summative assessment literacy development so they can be better prepared for the challenges of summatively assessing students.

Another implication of this study is that beginner teachers should be made aware that development is complex and variable, in order to understand that teachers follow idiosyncratic trajectories in their summative assessment literacy development. In this way beginner teachers gain the perspective that they are not unusual if they are not mirroring the learning patterns of their colleagues. It is

important that they accept that variability exists, and then focus on areas where they need to develop knowledge and understanding. In order to make informed choices about their learning as teachers, beginner teachers can be encouraged to be purposeful in self-review. The SALRubric could be used for this purpose.

Beginner teachers would be advised to carefully consider the amplifiers and filters that are operating for them and how several of these can interplay on multiple levels causing them to prioritise factors in assessment decisions differently at different times. They should aim for deliberate and purposeful development of assessment literacy over time rather than expecting that they will learn a narrow range of skills in ITE that will suffice for their summative assessment practice. Within this approach they could, for example, consider the stresses of emotional and ethical work they will be doing involving summative assessment, but call on a breadth of knowledge as they make decisions.

9.3 Limitations

It is important to consider the limitations of this study alongside the conclusions and implications drawn from it. This was a small qualitative study involving eight beginner secondary science teachers who completed a one year ITE programme at one university in New Zealand. Its aim was to give rich, in-depth accounts of the summative assessment literacy development for these teachers. A major issue for this sort of study is generalisability, as findings cannot necessarily be generalised from a qualitative study to apply to a whole population, or even to other ITE programmes. Nonetheless, readers of the findings of this study who are interested in the development of summative assessment literacy in beginner teachers are able to consider the findings and whether they may or may not be transferred to their own contexts. To aid decisions about transferability this thesis has included rich descriptions of the beginner teachers, context and process of analysis as well as appropriate quotations; such rich thick descriptions will help readers decide what findings can be transferred (Bryman, 2001; Geertz, 1973; Lincoln & Guba, 1985).

The beginner teachers involved in this study were a group of eight science teachers. They were volunteers, and their willingness to be involved in the study may reflect something of their particular interest and strengths. Although the group of beginner teachers were a whole class from one university programme,

their involvement was not designed to be representative of the population of beginner teachers in New Zealand. The beginner teachers were mostly young, with only one teacher aged over 30 years old, so most had experienced NCEA themselves as students. It is not known whether this age distribution may have affected their responses. There were three male and five female beginner teachers, which does represent the usual distribution in secondary science teachers. The views and perspectives of the beginner teachers were variable as were their professional and personal backgrounds. The variability in their responses could also be assumed to be found in the wider population of beginner teachers. This study focused on the development of teachers in an 18 month window of their ITE and first employment as teachers. Although there were considerable identified influences on their conceptions and practice of summative assessment in this time frame, a limitation is that other influences may have been overlooked or underplayed in the study. For example, in interviews, teachers may have overlooked influences that were in operation but not specifically identified at that time. Much of the data gathered in this study was self-report data, which means its quality was dependent on the honesty of the beginner teachers in their interview responses, which is a limitation. However, the use of artefacts provided a means for triangulation of data.

Abell and Siegel's (2011) model for science teacher assessment literacy, which provided a framework for this study, places teachers views of student learning in the core. However, I have not explored views of learning in the context of this study. The data that was gathered focused on teachers' assessment knowledge and decisions, and although some beginner teachers did talk about how they understood learning to occur, I felt there was insufficient data to make a substantiated judgement on teachers' views of learning.

School students are directly affected by beginner teachers' summative assessment decisions, and students' views may have provided alternative perspectives on summative assessment practice and decision making, so it may have been worthwhile to interview the students with whom the beginner teachers were working. However, it was outside the scope of the study to interview students, or to observe teachers working in their classes. This was a constraint in the research design as the link between beginner teachers' comments and their practice was

limited to what artefacts they provided to illustrate their summative assessment literacy. Classroom observation may have provided evidence of whether their summative assessment practice did reflect their reported practice. This was not possible, however, because of the constraints of time, while the size of this study precluded further data collection.

The questionnaire used in this study came from a larger study focused on primary preservice teachers, and it was not redesigned for use in this study. The questionnaire was broader than this study's focus on summative assessment as it was interested in teachers' development in formative assessment. However, the questionnaire did contribute to the study in that it was used to gain background information and occasionally questionnaire responses were used as a springboard to probe the ideas of individual beginner teachers during interviews.

The SALRubric was designed as a tool to track summative assessment literacy development over time, but the dimensions should not be seen as inclusive of all aspects of summative assessment literacy. The development of the SALRubric was limited by what was found in the literature and what was found through the analysis of data from this study. This means it was appropriate for use in this study. However, other dimensions of summative assessment literacy may exist, in particular in the category: Understanding the context for assessment. This category may vary considerably depending on the education system in which teachers work, so care needs to be taken when applying the SALRubric in other contexts.

A further limitation of this study was the method of data analysis. Because most analysis was conducted solely by the researcher it has to be accepted that researcher views are integral to the interpretation of the data. This is an accepted aspect of interpretivist research. The trustworthiness of findings has been strengthened through a process where independent researchers were asked to check the codes of interviews and a subset of judgements using the SALRubric on selections of data, to confirm the accuracy of decisions. Considerable use has been made of direct quotations from the beginner teachers so that readers can decide for themselves whether they agree with the interpretations presented. In this study a number of strategies were used to ensure the study is credible: triangulation,

prolonged engagement, member checks and peer validation. Finally, through this process of undertaking this research, I practiced reflexivity by regularly reflecting on my own position as a researcher and the way I engaged with all aspects of the data analysis. By doing this I worked to represent the beginner teachers' ideas, practices and development as accurately as possible.

9.4 Suggestions for further research

A number of avenues of useful research could be pursued, building on what has been achieved in this study.

The beginner science teachers involved in this study developed summative assessment literacy over time in idiosyncratic ways. It would be useful to broaden the study to investigate whether what was found would be more widely applicable for teachers of all secondary curriculum subjects, as well to explore summative assessment literacy as it applies to primary settings in the New Zealand context. Further investigation regarding the dimensions of summative assessment literacy where little or slow shift was found for beginner teachers would be another useful area for further research. Linked to this was teachers' lack of confidence in summative assessment task design, which is a very important element of teaching in the New Zealand context. Research that investigated how to deliver effective ITE and inservice professional development which fosters the development of both knowledge and confidence in this area would prove helpful.

Related to this it would be worth investigating whether the SALRubric was more widely applicable, or whether new or different dimensions emerged. International comparisons could also be developed to probe the differences evident in teachers' summative assessment literacy in different cultural and policy environments; for example, to study and compare the effects of a more centralised summative assessment system on teacher summative assessment literacy.

With respect to assessment in science, it was unanticipated that the beginner teachers did not communicate an understanding of the nature of science when they discussed what to assess. Additionally, although they showed they could identify major science ideas through CoRe design (Loughran et al., 2006) they did not consider the relevance of prioritising these key science concepts in summative

assessment. Further research to investigate how this could be remedied could be useful to inform ITE programmes.

The research design used in this longitudinal study (involving a combination of survey, interviews and artefacts) was effective in exploring teacher development over time. This study was unable to include the observation of summative assessment practice *in situ* in classrooms, which meant that only interviews and the selected artefacts provided by teachers were able to be used as evidence of summative assessment practice. A study which includes observations and the full spectrum of all summative assessments used may paint a more complete picture and would be worthwhile. It would be interesting to use similar or enhanced research design to track teacher development for even longer periods of time in order to gain a more comprehensive understanding of teacher development through a number of years of their careers, to note what changes occur as experience is gained.

With respect to ITE, this study has established that there were very specific identifiable learning activities that beginner teachers said made a big difference to their learning and confidence regarding summative assessment practice. Further investigation into what makes an ITE learning activity particularly useful for teacher development related to assessment would contribute to further understanding in this area. Associated with this could be the refinement of SALRubric as a tool that teachers could use for appraising their own progress, and that ITE lecturers could use together with preservice teachers in their courses.

The SALRubric itself has been informed by the literature in conjunction with the findings of this quantitative study. Although measures were put in place to ensure consistency in its application, research that results in further refinement of the tool would be useful in order to extend and confirm its applicability for teachers and ITE lecturers across different educational contexts.

9.5 Conclusion

Moss (2013) contends that "assessment is unquestionably one of the teacher's most complex and important tasks" (p. 235). Beginner teachers must engage in this task and this requires development of their summative assessment literacy.

This study extends past research and contributes to understandings of summative assessment literacy and how it develops. The complex amalgam of assessment knowledges together with the effects of amplifiers and filters leads to idiosyncratic development of summative assessment literacy for beginner teachers. The SALRubric has potential as a tool that would inform and guide teachers' summative assessment literacy development and lead to programmes that better prepare teachers to be more effective in meeting the challenges inherent in this most complex and important task.

References

Abd-El-Khalick, F. (2014). The evolving landscape related to assessment of Nature of Science. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of research on science education* (Vol. 2, pp. 621–650). New York, NY: Routledge.

Abell, S. K. (2007). Research on science teacher knowledge. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 1105–1150). Mahwah, NJ: Lawrence Erlbaum Associates.

Abell, S. K., & Siegel, M. A. (2011). Assessment literacy: What science teachers need to know and be able to do. In D. Corrigan, J. Dillon, & R. Gunstone (Eds.), *The professional knowledge base of science teaching* (pp. 205–221). Dordrecht, The Netherlands: Springer.

Abrahams, I., Reiss, M. J., & Sharpe, R. M. (2013). The assessment of practical work in school science. *Studies in Science Education*, *49*(2), 209–251. https://doi.org/10.1080/03057267.2013.858496

Absolum, M. (2006). *Clarity in the classroom*. Auckland, New Zealand: Hachette Livre.

Absolum, M., Flockton, L., Hattie, J., Hipkins, R., & Reid, I. (2009). Directions for assessment in New Zealand: Developing students' assessment capabilities. Retrieved 16 June 2012, from http://assessment.tki.org.nz/Assessment-in-the-classroom/Directions-for-assessment-in-New-Zealand-DANZ-report

Adie, L. E. (2011). An investigation into online moderation. *Assessment Matters*, *3*, 5–27.

Adie, L. E., Klenowski, V., & Wyatt-Smith, C. (2012). Towards an understanding of teacher judgement in the context of social moderation. *Educational Review*, *64*(2), 223–240. https://doi.org/10.1080/00131911.2011.598919

Adoniou, M. (2013). Preparing teachers: The importance of connecting contexts in teacher education. *Australian Journal of Teacher Education*, *38*(8). https://doi.org/10.14221/ajte.2013v38n8.7

Aitken, J. L., & Mildon, D. (1992). Teacher education and the developing teacher: The role of personal knowledge. In M. Fullan & A. Hargreaves (Eds.), *Teacher development and educational change* (pp. 10–35). London, England: Falmer.

Alexander, P. A., Schallert, D. L., & Hare, V. C. (1991). Coming to terms: How researchers in learning and literacy talk about knowledge. *Review of Educational Research*, *61*(3), 315–343. https://doi.org/10.3102/00346543061003315

Alison, J. (2005). *Teachers talk about NCEA: Research report on focus groups with secondary teachers* (pp. 1–167). Wellington, New Zealand: PPTA.

Alkharusi, H., Kazem, A. M., & Al-Musawai, A. (2011). Knowledge, skills, and attitudes of preservice and inservice teachers in educational measurement. *Asia-Pacific Journal of Teacher Education*, *39*(2), 113–123. https://doi.org/10.1080/1359866X.2011.560649

Allen, J. M. (2009). Valuing practice over theory: How beginning teachers re-orient their practice in the transition from the university to the workplace. *Teaching and Teacher Education*, *25*(5), 647–654. https://doi.org/10.1016/j.tate.2008.11.011

Alvesson, M. (2003). Beyond neopositivisits, romantics and localists: a reflective approach to interviews in organizational research. *Academy of Management Review*, *28*(1), 13-33.

Amrein-Beardsley, A., Berliner, D. C., & Rideau, S. (2010). Cheating in the first, second, and third degree: Educators' responses to high-stakes testing. *Educational Policy Analysis Archives*, *18*(14), 1–32. http://dx.doi.org/10.14507/epaa.v18n14.2010

Anderson, C. S. (2007). Perspectives on science teaching. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 3–30). New York, NY: Routledge.

Angrosino, M., & Rosenburg, J. (2011). Observations on observation: Continuities and challenges. In N. Denzin & Y. Lincoln (Eds.), *Qualitative research* (4th ed.). Washington, DC: Sage Publications.

Anthony, G., & Kane, R. (2008). *Making a difference: The role of initial teacher education and induction in the preparation of secondary teachers* (pp. 1–84). Wellington, New Zealand: Teaching and Learning Research Initiative.

Anttila, H., Pyhältö, K., Soini, T., & Pietarinen, J. (2016). How does it feel to become a teacher? Emotions in teacher education. *Social Psychology of Education*, *19*(3), 451–473. https://doi.org/10.1007/s11218-016-9335-0

Arzi, H. J., & White, R. T. (2008). Change in teachers' knowledge of subject matter: A 17-year longitudinal study. *Science Education*, *92*(2), 221–251.

Askew, M. (2013). Issues in teaching for and assessment of creativity in mathematics and science. In D. Corrigan, R. Gunstone, & A. Jones (Eds.), *Valuing assessment in science education: Pedagogy, curriculum, policy*. Dordrecht, The Netherlands: Springer.

Assessment Reform Group. (2006). *The role of teachers in assessment of learning*. London, England: Institute of Education, University of London.

Athanases, S. Z., Bennett, L. H., & Wahleithner, J. M. (2013). Fostering data literacy through pre-service teacher inquiry in English language arts. *The Teacher Educator*, *48*(1), 8–28. http://dx.doi.org/10.1080/08878730.2012.740151

Austin, L. G. (2015). *The decision making process leading to curriculum innovation in medium sized New Zealand secondary schools since the introduction of NCEA* (Master of Education thesis, Victoria University of

Wellington, Wellington, New Zealand). Retrieved from
http://hdl.handle.net/10063/4573

Aydin, S., Demirdogen, B., Akin, F. N., Uzuntirkaki-Kondakci, E., & Tarkin, A. (2015). The nature and development of interaction among components of pedagogical content knowledge in practicum. *Teaching and Teacher Education*, *46*, 37–50. https://doi.org/http://dx.doi.org/10.1016/j.tate.2014.10.008

Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., & van der Vleuten, C. P. M. (2007). Evaluating assessment quality in competence-based education: A qualitative comparison of two frameworks. *Educational Research Review*, *2*(2), 114–129. https://doi.org/10.1016/j.edurev.2007.06.001

Bakx, A., Baartman, L., & van Schilt-Mol, T. (2014). Development and evaluation of a summative assessment program for senior teacher competence. *Studies in Educational Evaluation*, *40*, 50–62. https://doi.org/10.1016/j.stueduc.2013.11.004

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching what makes it special? *Journal of Teacher Education*, *59*(5), 389–407. https://doi.org/10.1177/0022487108324554

Ball, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy*, *18*(2), 215–228. https://doi.org/10.1080/0268093022000043065

Banister, P., Burman, E., Parker, I., Taylor, M., & Tindall, C. (1994). *Qualitative methods in psychology*. Buckingham, PA: Open University Press.

Barrett, J. (2008). *The impact of the NCEA on teacher collegiality* (Thesis, Master of Educational Leadership, The University of Waikato, Hamilton, New Zealand. Retrieved from http://hdl.handle.net/10289/2270

Beattie, M. (1995). *Constructing professional knowledge in teaching: A narrative of change and development*. New York, NY: Teachers College Press.

Bell, B. (2007). Classroom assessment of science learning. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education*. London, England: Lawrence Erlbaum Associates.

Bell, B. (2011). *Theorising teaching in secondary classrooms: Understanding our practice from a sociocultural perspective*. New York, NY: Routledge.

Bell, B., & Cowie, B. (2001). *Formative assessment and science education*. Dordrecht, The Netherlands: Kluwer Academic.

Bell, B., & Gilbert, J. (1994). Teacher development as professional, personal, and social development. *Teaching and Teacher Education*, *10*(5), 483–497. https://doi.org/10.1016/0742-051X(94)90002-7

Bellocchi, A., Ritchie, S. M., Tobin, K., King, D., Sandhu, M., & Henderson, S. (2014). Emotional climate and high quality learning experiences in science teacher education. *Journal of Research in Science Teaching*, *51*(10), 1301–1325. https://doi.org/10.1002/tea.21170

Bennett, R. E., Gottesman, R. L., Rock, D. A., & Cerullo, F. (1993). Influence of behaviour perceptions and gender on teachers' judgements of students' academic skill. *Journal of Educational Psychology*, *85*, 347–356. https://doi.org/http://dx.doi.org/10.1037/0022-0663.85.2.347

Berliner, D. C. (2001). Improving the quality of the teaching force: A conversation with David C. Berliner. *Educational Leadership*, *58*(8), 6–11.

Berryman, M. (2015). A culturally responsive framework for social justice. In K. Esposito, & A. Normore (Eds.), *Inclusive practices and social justice for special populations in urban settings* (pp. 143–164). Charlotte, NC: Information Age Pub.

Bianchini, J. A. (2012). Teaching while still learning to teach: Beginning science teachers' views, experiences, and classroom practices. In B. J. Fraser, K. G. Tobin, & C. J. McRobbie (Eds.), *Second international handbook of science education* (Vol. 24, pp. 389–400). Dordrecht, The Netherlands: Springer.

Billett, S., Fenwick, T. J., & Somerville, M. (Eds.). (2006). *Work, subjectivity and learning understanding learning through working life*. Dordrecht, The Netherlands: Springer.

Bishop, R., & Berryman, M. (2009). The Te Kotahitanga effective teaching profile. *Set: Research Information for Teachers*, *2*, 27–33.

Black, P. (1998). *Testing: Friend or foe? Theory and practice of assessment and testing*. London, England: Falmer.

Black, P. (2001). *Report to the Qualifications Development Group (QDG) Ministry of Education, New Zealand on the proposals for the development of the NCEA*. Wellington, New Zealand: Ministry of Education.

Black, P. (2003). The importance of everyday assessment. In J. M. Atkin, & J. E. Coffey (Eds.), *Everyday assessment in the science classroom* (pp. 1–12). Arlington, VA: National Science Teachers Association.

Black, P. (2013a). Pedagogy in theory and in practice: Formative and summative assessments in classrooms and in systems. In D. Corrigan, R. Gunstone, & A. Jones (Eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy* (pp. 207–229). Dordrecht, The Netherlands: Springer.

Black, P. (2013b). Formative and summative aspects of assessment: Theoretical and research foundations in the context of pedagogy. In J.H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 167–178). Thousand Oaks, CA: Sage Publications.

Black, P. & Atkin, J.M. (1996). *Changing the subject: Innovations in science mathematics and technology education*. Paris, France: OECD

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. https://doi.org/10.1080/09695941003696016

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. https://doi.org/10.1080/0969594X.2011.557020

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Maidenhead, England: Open University Press.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *The Phi Delta Kappan*, *80*(2), 139–148. https://doi.org/10.1177/003172171009200119

Black, P., & Wiliam, D. (2003). 'In praise of educational research': Formative assessment. *British Educational Research Journal, 29*(5), 623-637. DOI: 10.1080/0141192032000133721

Black, P., & Wiliam, D. (2005). Lessons from around the world: How policies, politics and cultures constrain and afford assessment practices. *Curriculum Journal*, *16*(2), 249–261. https://doi.org/10.1080/09585170500136218

Black, P., & Wiliam, D. (2006). Assessment for learning in the classroom. In J. Gardner (Ed.), *Assessment and learning*. (pp. 11-32). London, England: Sage Publications.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Bol, L., & Strage, A. (1996). The contradiction between teachers' instructional goals and their assessment practices in high school biology courses. *Science Education*, *80*(2), 145–163. https://doi.org/10.1002/(SICI)1098-237X(199604)80:2<145::AID-SCE2>3.0.CO;2-G

Bonner, S.M., & Chen, P.P. (2009). Teacher candidates' perceptions about grading and constructivist teaching. *Educational Assessment, 14*(2), 57-77.

Booth, B., Hill, M. F., & Dixon, H. R. (2014). The assessment-capable teacher: Are we all on the same page? *Assessment Matters*, *6*, 137–157. https://doi.org/http://dx.doi.org/10.18296/am

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, *33*(8), 3–15. https://doi.org/10.3102/0013189x033008003

Borowski, A., Carlson, J., Fischer, H. E., Henze, I., & Gess-Newsome, J. (2012). Different models and methods to measure teachers' pedagogical content knowledge. In C. Bruguiere, A. Tiberghien, & P. Clement (Eds.), *EBook proceedings of the ESERA 2011 conference - Science learning and*

citizenship. (pp. 25-36). Lyon, France: European Science Education Research Association.

Boyd, B., & Hayward, L. (2010). *Exploring assessment for accountability*. Glasgow, Scotland: Universities of Strathclyde and Glasgow.

Bransford, J. D., Brown, A. L., & Cocking, R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Brenner, M. E. (2006). Interviewing in educational research. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 357–370). Washington, DC: Lawrence Erlbaum Associates.

Britton, E. D., & Schneider, S. A. (2007). Large-scale assessments in science education. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education*. (pp.1007-1042). London, England: Lawrence Erlbaum Associates.

Broadfoot, P. (2012). *Assessment, schools and society* (Vol. 35). Oxon, England: Routledge.

Brookhart, S. M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice*, *10*(1), 35–36. https://doi.org/10.1111/j.1745-3992.1991.tb00182.x

Brookhart, S. M. (1997). A theoretical framework for the role of classroom assessment in motivating student effort and achievement. *Applied Measurement in Education*, *10*(2), 161–180. https://doi.org/http://dx.doi.org/10.1207/s15324818ame1002_4

Brookhart, S. M. (1999). *The art and science of classroom assessment: The missing part of pedagogy* (ASHE-ERIC Higher Education Report No. Vol. 27, No. 1). Washington, DC: The George Washington University, Graduate School of Education and Human Development.

Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement, Issues and Practice*, *22*(4), 5–12.

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, *30*(1), 3–12. https://doi.org/http://dx.doi.org/10.1111/j.1745-3992.2010.00195.x

Brookhart, S. M. (2013a). Grading. In J.H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 257–271). Thousand Oaks, CA: Sage Publications.

Brookhart, S. M. (2013b). Research on classroom summative assessment. In J.H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 235–256). Thousand Oaks, CA: Sage Publications.

Brookhart, S. M. (2013c). Classroom assessment in the context of motivation theory and research. In J.H. McMillan (Ed.), *Sage handbook of research*

on classroom assessment (pp. 35–54). Thousand Oaks, CA: Sage Publications.

Brown, C. R., Moor, J. L., Silkstone, B. E., & Botton, C. (1998). An evaluation of two different methods of assessing independent investigations in an operational pre-university level examination in biology in England. *Studies in Educational Evaluation*, *24*, 87–98. https://doi.org/http://dx.doi.org/10.1016/s0191-491x(98)00005-4

Brown, G. T. L. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York, NY: Nova Science Publishers.

Brown, G. T. L. (2011). Teachers' conceptions of assessment: Comparing primary and secondary teachers in New Zealand. *Assessment Matters*, *3*, 45–70.

Brown, G. T. L., Irving, E., & Keegan, P. (2008). *An introduction to educational assessment, measurement and evaluation* (2nd ed.). Auckland, New Zealand: Pearson Education NZ.

Brown, G. T. L., Irving, S. E., & Keegan, P. (2014). *An introduction to educational assessment: Measurement & evaluation* (3rd ed.). Auckland, New Zealand: Dunmore Press.

Brown, G. T. L., Irving, S. E., Peterson, E. R., & Hirschfeld, G. H. F. (2009). Use of interactive–informal assessment practices: New Zealand secondary students' conceptions of assessment. *Learning and Instruction*, *19*(2), 97–111. https://doi.org/10.1016/j.learninstruc.2008.02.003

Brown, G. T. L., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and Teacher Education*, *27*, 210–220. https://doi.org/http://dx.doi.org/10.1016/j.tate.2010.08.003

Brown, G. T. L., & Michaelides, M. P. (2011). Ecological rationality in teachers' conceptions of assessment across samples from Cyprus and New Zealand. *European Journal of Psychology of Education*, *26*(3), 319–337. https://doi.org/10.1007/s10212-010-0052-3

Brown, G. T. L., & Wang, Z. (2013). Illustrating assessment: How Hong Kong university students conceive of the purposes of assessment. *Studies in Higher Education*, *38*(7), 1037–1057. https://doi.org/10.1080/03075079.2011.616955

Bryman, A. (2001). *Social research methods*. Oxford, England: Oxford University Press.

Buck, G., & Trauth-Nare, A. (2009). Preparing teachers to make the formative assessment process integral to science teaching and learning. *Journal of Science Teacher Education*, *20*(5), 475–494. https://doi.org/http://dx.doi.org/10.1007/s10972-009-9142-y

Bullock, K., Bishop, K., Martin, S., & Reid, A. (2002). Learning from coursework in English and geography. *Cambridge Journal of Education*, *32*(3), 325–340. https://doi.org/10.1080/0305764022000024195a

Bullough, R. V. (1992). Beginning teacher curriculum decision making, personal teaching metaphors, and teacher education. *Teaching and Teacher Education*, *8*(3), 239–252. https://doi.org/http://dx.doi.org/10.1016/0742-051x(92)90023-v

Bullough, R. V. (2011). Ethical and moral matters in teaching and teacher education. *Teaching and Teacher Education*, *27*(1), 21–28. https://doi.org/10.1016/j.tate.2010.09.007

Burrell, G. & Morgan, G. (1979). Sociological paradigms and organisational analysis: Elements of the sociology of corporate life. London, England: Heinnemann.

Buxton, C. A., Salinas, A., Mahotiere, M., Lee, O., & Secada, W. G. (2015). Fourth-grade emergent bilingual learners' scientific reasoning complexity, controlled experiment practices, and content knowledge when discussing school, home, and play contexts. *Teachers College Record*, *117*(2), 1–36.

Cameron, M., & Baker, R. (2004). *Research on initial teacher education in New Zealand: 1993-2004*. Wellington, New Zealand: NZCER Press.

Cameron, S., Mulholland, J., & Branson, C. (2013). Professional learning in the lives of teachers: Towards a new framework for conceptualising teacher learning. *Asia-Pacific Journal of Teacher Education*, *41*(4), 377–397. https://doi.org/10.1080/1359866X.2013.838620

Campbell, E. (2003). *The ethical teacher*. Maidenhead, England: Open University Press.

Campbell, S. (2008). *Characteristics of reflection: Beginning science teachers' video-stimulated discussion of their lessons* (Masters thesis, University of London, England).

Carless, D. (2005). Prospects for the implementation of assessment for learning. *Assessment in Education: Principles, Policy & Practice*, *12*(1), 39–54. https://doi.org/10.1080/0969594042000333904

Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International*, *44*(1), 57–66. https://doi.org/10.1080/14703290601081332

Carless, D., & Lam, R. (2014). The examined life: Perspectives of lower primary school students in Hong Kong. *Education 3-13*, *42*(3), 313–329. https://doi.org/10.1080/03004279.2012.689988

Carr, D. (1986). *Time, history and narrative*. Bloomington, IN: Indiana University Press.

Carr, M., McGee, C., Jones, A., McKinley, E., Bell, B., Barr, H., & Simpson, T. (2005). *The effects of curricula and assessment on pedagogical approaches and on educational outcomes*. Wellington, New Zealand: Ministry of Education.

Chamberlain, S. (2013). Communication strategies for enhancing qualification users' understanding of educational assessment: Recommendations from

other public interest fields. *Oxford Review of Education*, *39*(1), 114–127. https://doi.org/10.1080/03054985.2013.764757

Chant, R. H., Heafner, T. L., & Bennett, K. R. (2004). Connecting personal theorizing and action research in preservice teacher development. *Teacher Education Quarterly*, *31*(3), 25–42.

Chazan, D., Brantlinger, A., Clark, L. M., & Edwards, A. R. (2013). What mathematics education might learn from the work of well-respected African American mathematics teachers in urban schools. *Teachers College Record*, *115*(2), 1–40.

Chen, P. P. (2005). Teacher candidates' literacy in assessment. *Academic Exchange Quarterly*, *65*(5), 62–66.

Cheng, M. M. H., Tang, S. Y. F., & Cheng, A. Y. N. (2012). Practicalising theoretical knowledge in student teachers' professional learning in initial teacher education. *Teaching and Teacher Education*, *28*(6), 781–790. https://doi.org/10.1016/j.tate.2012.02.008

Chong, S., Ling, L. E., & Chuan, G. K. (2011). Developing student teachers' professional identities - An exploratory study. *International Education Studies*, *4*(1), 30–38. https://doi.org/10.5539/ies.v4n1p30

Christoforidou, M., Kyriakides, L., Antoniou, P., & Creemers, B. P. M. (2014). Searching for stages of teacher's skills in assessment. *Studies in Educational Evaluation*, *40*, 1–11. https://doi.org/10.1016/j.stueduc.2013.11.006

Clark, C. M., & Petersen, P. L. (1986). Teachers' thought processes. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 255–296). New York, NY: Macmillan.

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, *18*(8), 947–967. https://doi.org/10.1016/S0742-051X(02)00053-7

Clarke, S., Timperley, H., & Hattie, J. (2003). *Unlocking formative assessment*. Auckland, New Zealand: Hodder Moa Beckett.

Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher*, *23*(7), 13–20. https://doi.org/10.2307/1176934

Cobb, P. (2007). Putting philosophy to work: Coping with multiple theoretical perspectives. In F. K. Lester (Ed.), *Second handbook on research on mathematics teaching and learning* (Vol. 1, pp. 3–38). Reston, VA: National Council of Teachers of Mathematics.

Cobb, P., & Bauersfeld, H. (Eds.). (1995). *The emergence of mathematical meaning: Interaction in classroom cultures*. Hillsdale, NJ: L. Erlbaum Associates.

Cobb, P., Jaworski, B., & Presmeg, N. (1996). Emergent and sociocultural views of mathematical activity. In P. Nesher, L. P. Steffe, P. Cobb, G. Goldin, &

B. Greer (Eds.), *Theories of mathematical learning* (pp. 3–20). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Cobb, P., & Yackel, E. (1996). Constructivist, emergent, and sociocultural perspectives in the context of developmental research. *Educational Psychologist*, *31*(3), 175–190. https://doi.org/10.1207/s15326985ep3103&4_3

Cochran-Smith, M. (2004). Ask a different question, get a different answer: The research base for teacher education. *Journal of Teacher Education*, *55*(2), 111–115. https://doi.org/10.1177/0022487104262971

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education* (5th ed.). London, England: RoutledgeFalmer.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). New York, NY: Routledge.

Coll, R. K., & Chapman, R. (2000). Choices of methodology for cooperative researchers. *Asia-Pacific Journal of Cooperative Education*, *1*, 1–8.

Colnerud, G. (1997). Ethical conflicts in teaching. *Teaching and Teacher Education*, *13*(6), 627–635. https://doi.org/10.1016/S0742-051X(97)80005-4

Colnerud, G. (2006). Teacher ethics as a research problem: Syntheses achieved and new issues. *Teachers and Teaching*, *12*(3), 365–385. https://doi.org/10.1080/13450600500467704

Connell, R. (2013). The neoliberal cascade and education: An essay on the market agenda and its consequences. *Critical Studies in Education*, *54*(2), 99–112. https://doi.org/10.1080/17508487.2013.776990

Connelly, F. M., & Clandinin, D. J. (1995). Teachers' professional knowledge landscapes: Secret, sacred, and cover stories. In D. J. Clandinin & F. M. Connelly (Eds.), *Teachers' professional knowledge landscapes* (pp. 3–15). New York, NY: Teachers College Press.

Cooper, B., & Edwards, F. (2013, April). Tensions in developing assessment capability with and for secondary student teachers. Presented at the American Educational Research Association, San Francisco, CA.

Corrigan, D., Buntting, C., Jones, A., & Gunstone, R. (2013). Valuing assessment in science education: An introductory framework. In D. Corrigan, R. Gunstone, & A. Jones (Eds.), *Valuing assessment in science education: Pedagogy, curriculum, policy* (pp. 1–10). Dordrecht, The Netherlands: Springer.

Corrigan, D., & Cooper, R. (2013). The influence of assessment on moderating science teachers' beliefs and practices. In D. Corrigan, R. Gunstone, & A. Jones (Eds.), *Valuing assessment in science education: Pedagogy, curriculum, policy* (pp. 153–168). Dordrecht, The Netherlands: Springer.

Cowan, E. M. (2009). Implementing formative assessment: Student teachers' experiences on placements. *Teacher Development: An International*

Journal of Teachers' Professional Development, *13*(1), 71–84.
https://doi.org/http://dx.doi.org/10.1080/13664530902858519

Cowie, B. (2013). Assessment in the science classroom: Priorities, practices, and prospects. In J. H. McMillan (Ed.), *Sage handbook on research on classroom assessment* (pp. 473–488). Thousand Oaks, CA: Sage Publications.

Cowie, B. (2015). Equity, ethics and engagement: Principles for quality formative assessment in primary science classrooms. In C. Milne, K. G. Tobin, & D. DeGennaro (Eds.), *Sociocultural studies and implications for science education: The experiential and the virtual* (pp. 117–134). Dordrecht, The Netherlands: Springer.

Creswell, J. W. (2012). *Educational research: Planning, conducting and evaluating quantitative and qualitative research* (4th ed.). Boston, MA: Pearson.

Cronin-Jones, L. L. (1991). Science teachers' beliefs and their influence on curriculum implementation - two case studies. *Journal of Research in Science Teaching*, *28*(3), 235–250. https://doi.org/http://dx.doi.org/10.1002/tea.3660280305

Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, *58*(4), 438–481. https://doi.org/http://dx.doi.org/10.3102/00346543058004438

Crooks, T. (2002). Educational assessment in New Zealand schools. *Assessment in Education: Principles, Policy & Practice, 9*(2), 237-253. 10.1080/0969594022000001959

Crooks, T. (2011). Assessment for learning in the accountability era: New Zealand. *Studies in Educational Evaluation*, *37*(1), 71–77. https://doi.org/http://dx.doi.org/10.1016/j.stueduc.2011.03.002

Crooks, T. J., Kane, M. T., & Cohen, A. S. (1996). Threats to the valid use of assessments. *Assessment in Education: Principles, Policy & Practice*, *3*(3), 265–286. https://doi.org/10.1080/0969594960030302

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process.* Crows Nest, Australia: Allen & Unwin.

Daly, N. (2008). Not empty vessels: New Zealand pre-service additional language teacher identity. *Waikato Journal of Education*, *14*, 5–13. https://doi.org/http://dx.doi.org/10.15663/wje.v14i1.211

Damasio, A. R. (2005). The neurobiological grounding of human values. In J. P. Changeux, A. R. Damasio, W. Singer, & Y. Christen (Eds.), *Neurobiology of human values*. London, England: Springer.

Danielewicz, J. (2001). *Teaching selves, identity, pedagogy, and teacher development*. New York, NY: State University of New York Press.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, *8*(1).

Retrieved from
http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ605912

Darling-Hammond, L. (2006a). *Powerful teacher education*. San Francisco, CA:
Jossey-Bass.

Darling-Hammond, L. (2006b). Constructing 21st-century teacher Education.
Journal of Teacher Education, *57*(3), 300–314.
https://doi.org/10.1177/0022487105285962

Darling-Hammond, L. (2010). Teacher education and the American future.
Journal of Teacher Education, *6*(1–2), 35–47.
https://doi.org/http://dx.doi.org/10.1177/0022487109348024

Darling-Hammond, L., Hammerness, K., Grossman, P. L., Rust, F., & Shulman,
L. (2005). The design of teacher education programs. In L. Darling-
Hammond & J. D. Bransford (Eds.), *Preparing teachers for a changing
world: What teachers should learn and be able to do* (pp. 390–441). San
Francisco, CA: Jossey-Bass.

Darling-Hammond, L., Herman, J., Pellegrino, J., Abedi, J., Aber, J. L., Baker,
E., … Steele, C. M. (2013). *Criteria for high-quality assessment*. Stanford,
CA: Stanford Center for Opportunity Policy in Education.

Datnow, A., Park, V., & Wohlstetter, P. (2007). *Achieving with data. How-high
performing school systems use data to improve instruction for elementary
students.* San Francisco, CA: Center on Educational Governance
University of California.

Davies, A., & Hill, M. (2009). *Making classroom assessment work*. Wellington,
New Zealand: NZCER Press.

Davies, P. B., & Bansel, P. (2007). Neoliberalism and education. *International
Journal of Qualitative Studies in Education*, *20*(3), 247–259.
https://doi.org/10.1080/09518390701281751

Day, C., & Gu, Q. (2014). *Resilient teachers, resilient schools. Building and
sustaining quality in testing times.* London, England: Routledge.

DeLuca, C., Chavez, T., Bellara, A., & Cao, C. (2013). Pedagogies for preservice
assessment education: Supporting teacher candidates' assessment literacy
development. *The Teacher Educator*, *48*(2), 128–142.
https://doi.org/10.1080/08878730.2012.760024

DeLuca, C., Chavez, T., & Cao, C. (2013). Establishing a foundation for valid
teacher judgement on student learning: The role of pre-service assessment
education. *Assessment in Education: Principles, Policy & Practice*, *20*(1),
107–126. https://doi.org/10.1080/0969594X.2012.668870

DeLuca, C., & Johnson, S. (2017). Developing assessment capable teachers in this
age of accountability. *Assessment in Education: Principles, Policy &
Practice*, *24*(2), 121–126.
https://doi.org/10.1080/0969594X.2017.1297010

DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development:
Identifying gaps in teacher candidates' learning. *Assessment in Education:*

Principles, Policy & Practice, *17*(4), 419–438. https://doi.org/http://dx.doi.org/10.1080/0969594X.2010.516643

DeLuca, C., Klinger, D. A., Searle, M., & Shulha, L. M. (2010). Developing a curriculum for assessment education. *Assessment Matters*, *2*, 20–42.

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016a). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, *28*(3), 251–272. https://doi.org/10.1007/s11092-015-9233-6

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016b). Approaches to Classroom Assessment Inventory: A New Instrument to Support Teacher Assessment Literacy. *Educational Assessment*, *21*(4), 248-266. https://doi.org/10.1080/10627197.2016.1236677

DeLuca, C., Ogden, H., & Pero, E. (2015). Reconceptualizing elementary preservice teacher education: Examining an integrated-curriculum approach. *The New Educator*, *11*(3), 227–250. https://doi.org/10.1080/1547688X.2014.960986

Deneen, C. C., & Brown, G. T. L. (2016). The impact of conceptions of assessment on assessment literacy in a teacher education program. *Cogent Education*, *3*(1), 1-14. https://doi.org/10.1080/2331186X.2016.1225380

Denzin, N., & Lincoln, Y. S. (Eds.). (2011). *The Sage handbook of qualitative research* (4th ed.). Thousand Oaks, CA: Sage Publications.

Denzin, N., & Lincoln, Y.S. (Eds). (2013). Collecting and interpreting qualitative materials. Thousand Oaks, CA: Sage Publications.

Desimone, L. M., & Le Floch, K. C. (2004). Are we asking the right questions? Using cognitive interviews to improve surveys in education research. *Educational Evaluation and Policy Analysis*, *26*(1), 1–22.

Dixon, H. R., & Haigh, M. (2009). Changing mathematics teachers' conceptions of assessment and feedback. *Teacher Development: An International Journal of Teachers' Professional Development*, *13*(2), 173–186. https://doi.org/http://dx.doi.org/10.1080/13664530903044002

Dixon, H. R., Hawe, E., & Parr, J. (2011). Enacting assessment for learning: The beliefs practice nexus. *Assessment in Education: Principles, Policy & Practice*, *18*(4), 365–379. https://doi.org/10.1080/0969594X.2010.526587

Donovan, M. S., & Bransford, J. D. (Eds.). (2005). *How students learn: History, mathematics and science in the classroom.* Washington, DC: National Academies Press.

Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. New York, NY: Free Press.

Dreyfus, S. E. (2004). The five-stage model of adult skill acquisition. *Bulletin of Science, Technology & Society, 24*, 177–181. doi:10.1177/0270467604264992

Driver, R., Asoko, H., Leach, J., Mortimer, E., & Scott, P. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher*, *23*(7), 5–12.

Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education*, *32*, 268–291.

Earl, L. M. (2013). *Assessment as learning: Using classroom assessment to maximise student learning* (2nd ed.). Thousand Oaks, CA: Corwin.

Earl, L. M., & Katz, S. (2006). *Leading schools in a data-rich world. Harnessing data for school improvement.* Thousand Oaks, CA: Corwin.

Ecclestone, K., & Pryor, J. (2003). 'Learning careers' or 'assessment careers'? The impact of assessment systems on learning. *British Educational Research Journal*, *29*(4), 471–488. http://dx.doi.org/10.1080/01411920301849

Edelenbos, P., & Kubanek-German, A. (2004). Teacher assessment: The concept of 'diagnostic competence'. *Language Testing*, *21*(3), 259–283. https://doi.org/10.1191/0265532204lt284oa

Education Council New Zealand. (2015). Practising teacher criteria. Retrieved from http://educationcouncil.org.nz/content/practising-teacher-criteria

Edwards, A. (2009). Becoming a teacher. In H. Daniels, J. Porter, & H. Lauder (Eds.), *Educational theories, cultures and learning* (Vol. 1, pp. 153–164). London, England: Routledge.

Edwards, F. (2003). *Netting the maroro: An exploration of Cook Islands teachers' beliefs about language learning and teaching* (Master of Arts Thesis, Massey University, Palmerston North, New Zealand). Retrieved from http://hdl.handle.net/10179/11011#sthash.uADatJUm.dpuf

Edwards, F. (2013a). Assessing New Zealand high school science: Considerations for teachers' assessment literacy. *Asia-Pacific Forum on Science Learning and Teaching*, *14*(2), article 7.

Edwards, F. (2013b). Quality assessment by science teachers: Five focus areas. *Science Education International*, *24*(2), 212–226. http://hdl.handle.net/10289/8243

Edwards, F., & Cooper, B. (2012, August). Developing assessment capable secondary teachers. Paper presented at the European Association for Research on Learning and Instruction: Sig1, Brussels, Belgium.

Edwards, F. C. E., & Edwards, R. J. (2017). A story of culture and teaching: the complexity of teacher identity formation. *The Curriculum Journal*, *28*(2), 190-211. https://doi.org/10.1080/09585176.2016.1232200

Eisenhart, M. (2006). Representing qualitative data. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 567–581). Washington, DC: Lawrence Erlbaum Associates.

Elbaz, F. (1992). Hope, attentiveness, and caring for difference: The moral voice in teaching. *Teaching and Teacher Education*, *8*(5), 421–432. https://doi.org/10.1016/0742-051X(92)90047-7

Ell, F., Hill, M., & Grudnoff, L. (2012). Finding out more about teacher candidates' prior knowledge: Implications for teacher educators. *Asia-Pacific Journal of Teacher Education*, *40*(1), 55–65. https://doi.org/10.1080/1359866X.2011.643760

Elo, S., & Kyngas, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, *62*(1), 107–115. https://doi.org/10.1111/j.1365-2648.2007.04569.x

Eraut, M. (1994). *Developing professional knowledge and competence*. London, England: Falmer Press.

Eraut, M. (2009). How professionals learn through work. In N. Jackson (Ed.), *Learning to be a professional through higher education e-book.* (pp. 1-18). University of Surrey, England: Author House.

Eyers, G. (2014). *Preservice teachers' assessment learning: Change, development and growth* (Doctoral thesis, University of Auckland, New Zealand). Retrieved from http://hdl.handle.net/2292/23757

Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational Research*, *38*(1), 47.

Fastier, M. (2009). Managing NCEA assessment practice more effectively. *New Zealand Geographer*, *65*(2), 147–153. https://doi.org/10.1111/j.1745-7939.2009.01159.x

Flockton, L. (2012). *The development of the student assessment system in New Zealand.* (SABER - Systems approach for better education results No. 71517). Washington, DC: World Bank.

Flores, M. A. (2003, January). *Teacher learning in the workplace: Processes and influencing factors.* Paper presented at the Hawaii International Conference on Education, Hawaii. Retrieved from http://files.eric.ed.gov/fulltext/ED474177.pdf

Flores, M. A., & Day, C. (2006). Contexts which shape and reshape new teachers' identities: A multi-perspective study. *Teaching and Teacher Education*, *22*(2), 219–232. https://doi.org/http://dx.doi.org/10.1016/j.tate.2005.09.002

Forgasz, H. (2010). Streaming for mathematics in years 7–10 in Victoria: An issue of equity? *Mathematics Education Research Journal*, *22*(1), 57–90. https://doi.org/10.1007/BF03217559

Fried, L., Mansfield, C., & Dobozy, E. (2015). Teacher emotion research: Introducing a conceptual model to guide future research. *Issues in Educational Research*, *25*(4), 415–441.

Friedrichsen, P., & Berry, A. (2015). Science teacher PCK learning progressions. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining*

pedagogical content knowledge in science education (pp. 214–228). Thousand Oaks, CA: Routledge.

Furtak, E. M., Morrison, D., & Kroog, H. (2014). Investigating the link between learning progressions and classroom assessment. *Science Education, 98*, 640–673. doi:10.1002/sce.21122

Gale, J., Wind, S., Koval, J., Dagosta, J., Ryan, M., & Usselman, M. (2016). Simulation-based performance assessment: an innovative approach to exploring understanding of physical science concepts. *International Journal of Science Education*, *38*(14), 2284–2302. https://doi.org/10.1080/09500693.2016.1236298

Gardner, J. (Ed.). (2006). *Assessment and learning*. London, England: Sage Publications.

Gardner, J., Harlen, W., Hayward, L., & Stobart, G. (2011). Engaging and empowering teachers in innovative assessment practice. In R. Berry & B. Adamson (Eds.), *Assessment reform in education* (pp. 105–120). New York, NY: Springer.

Gardner, J., Harlen, W., Hayward, L., Stobart, G., & Montgomery, M. (2010). *Developing teacher assessment*. Maidenhead, England: Open University Press.

Gearhart, M., Nagashima, S., Pfotenhauer, J., Clark, S., Schwab, C., Vendlinski, T., … Bernbaum, D. J. (2006). Developing expertise with classroom assessment in K–12 science: learning to interpret student work. Interim findings from a 2-year study. *Educational Assessment*, *11*(3–4), 237–263. https://doi.org/10.1080/10627197.2006.9652990

Gearhart, M., & Osmundson, E. (2009). Assessment portfolios as opportunities for teacher learning. *Educational Assessment*, *14*(1), 1–24. https://doi.org/10.1080/10627190902816108

Gee, J. P. (2008). A sociocultural perspective on opportunity to learn. In P. A. Moss, D. C. Pullin, J. P. Gee, E. H. Haertel, & L. J. Young (Eds.), *Assessment, equity, and opportunity to learn*. New York, NY: Cambridge University Press.

Geertz, C. (1973). *The interpretation of cultures*. New York, NY: Perseus Books Group.

Gergen, M., & Gergen, K. (2000). Qualitative inquiry: Tensions and transformations. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (pp. 1025–1046). Thousand Oaks, CA: Sage Publications.

Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK: Results of the thinking from the PCK Summit. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 48–42). New York, NY: Routledge.

Gess-Newsome, J., & Lederman, N. G. (Eds.). (1999). *Examining pedagogical content knowledge: The construct and its implications for science education*. Boston, MA: Kluwer.

Gibbs, C. (1995). Turning a blind eye. Student teachers' concerns on teaching practice. *New Zealand Principal, November,* 13-17

Giles, D. (2010). Developing pathic sensibilities: A critical priority for teacher education programmes. *Teaching and Teacher Education*, *26*(8), 1151–1519. https://doi.org/http://dx.doi.org/10.1016/j.tate.2010.05.007

Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. London, England: Falmer Press.

Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research in Education*, *24*, 355–392. https://doi.org/10.2307/1167274

Gipps, C., Brown, M., McCallum, B., & McAlister, S. (1995). *Intuition or evidence?* Bristol, PA: Open University Press.

Goodnough, K., & Hung, W. (2009). Enhancing pedagogical content knowledge in elementary science. *Teaching Education*, *20*(3), 229–242. https://doi.org/10.1080/10476210802578921

Gottheiner, D. M., & Siegel, M. A. (2012). Experienced middle school science teachers' assessment literacy: Investigating knowledge of students' conceptions in genetics and ways to shape instruction. *Journal of Science Teacher Education*, *23*(5), 531–557. https://doi.org/10.1007/s10972-012-9278-z

Graham, J., Meyer, L., McKenzie, L., McClure, J., & Weir, K. F. (2010). Māori and Pacific secondary student and parent perspectives on achievement, motivation and NCEA. *Assessment Matters*, *2*, 132–157.

Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through preservice teacher education. *Teaching and Teacher Education*, *21*(6), 607–621. https://doi.org/10.1016/j.tate.2005.05.001

Grainger, P. R., & Adie, L. (2014). How do preservice teacher education students move from novice to expert assessors? *Australian Journal of Teacher Education, 39*(7), 89-105. http://dx.doi.org/10.14221/ajte.2014v39n7.9

Grant, C., & Agosto, V. (2008). Teacher capacity and social justice in teacher education. In M. Cochran-Smith, S. Feiman-Nemser, J. McIntyre, & K. Demers (Eds.), *Handbook of Research on Teacher Education: Enduring Questions and Changing Contexts* (pp. 175–200). Mahwah, NJ: Lawrence Erlbaum Publishers.

Graue, M., & Walsh, D. (1998). *Studying children in context: Theories, methods, and ethics*. London, England: Sage Publications.

Green, S. K., Johnson, R. L., Kim, D.-H., & Pope, N. S. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching and Teacher Education*, *23*(7), 999–1011. https://doi.org/10.1016/j.tate.2006.04.042

Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education*. New York, NY: Teachers College Press.

Grossman, P. L., Valencia, S. W., Evans, K., Thompson, C., Martin, S., & Place, N. (2000). Transitions into Teaching: Learning to Teach Writing in Teacher Education and Beyond. *Journal of Literacy Research*, *32*(4), 631–662. https://doi.org/10.1080/10862960009548098

Grudnoff, A. B. (2007). *Becoming a teacher: An investigation of the transition from student teacher to teacher* (PhD). University of Waikato, Hamilton, New Zealand.

Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation: Improving the effectiveness of evaluation results through responsive and naturalistic approaches.* San Francisco, CA: Jossey-Bass.

Guba, E. G., & Lincoln, Y. S. (2005). Paradigmatic controversies, contradictions, and emerging confluences. In N. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 191–215). Thousand Oaks, CA: Sage Publications.

Gustafson, B., Guilbert, S., & MacDonald, D. (2002). Beginning elementary science teachers: Developing professional knowledge during a limited mentoring experience. *Research in Science Education*, *32*(3), 281–302. https://doi.org/10.1023/A:1020809916037

Hagenauer, G., Hascher, T., & Volet, S. E. (2015). Teacher emotions in the classroom: associations with students' engagement, classroom discipline and the interpersonal teacher-student relationship. *European Journal of Psychology of Education*, *30*(4), 385–403. https://doi.org/10.1007/s10212-015-0250-0

Haggarty, L., & Postlethwaite, K. (2012). An exploration of changes in thinking in the transition from student teacher to newly qualified teacher. *Research Papers in Education*, *27*(2), 241–262. https://doi.org/10.1080/02671520903281609

Haigh, M. A., & Anthony, G. J. (2012). Induction and efficacy: A case study of New Zealand newly qualified secondary science teachers. *Journal of Science Teacher Education*, *23*(6), 651–671. https://doi.org/10.1007/s10972-012-9285-0

Haigh, M., Pinder, H., & McDonald, L. (2008). The practicum's contribution to students learning to teach. In C. Rubie-Davies & C. Rawlinson (Eds.), *Challenging thinking about teaching and learning* (pp. 251–268). New York, NY: Nova Science Publishers.

Hall, A. (2001, April). What ought I to do, all things considered? An approach to the exploration of ethical problems by teachers. Paper presented at the IIPE Conference, Key Centre for Ethics, Law, Justice and Governance, Griffith University, Brisbane, Australia.

Hargreaves, A. (1998). The emotional practice of teaching. *Teaching and Teacher Education*, *14*, 835–854.

Hargreaves, A. (2000). Mixed emotions: teachers' perceptions of their interactions with students. *Teaching and Teacher Education*, *16*(8), 811–826. https://doi.org/10.1016/S0742-051X(00)00028-7

Hargreaves, A. (2005). The emotions of teaching and educational change. In A. Hargreaves (Ed.), *Extending educational change: International handbook of educational change* (pp. 278–295). Dordrecht, The Netherlands: Springer.

Hargreaves, D. J. (1997). Student learning and assessment are inextricably linked. *European Journal of Engineering Education*, *22*(4), 401–409. https://doi.org/10.1080/03043799708923471

Harlen, W. (2004). *A systematic review of the evidence of the impact on students, teachers and the curriculum of the process of using assessment by teachers for summative purposes*. London, England: University of London

Harlen, W. (2005a). Trusting teachers' judgement: research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research Papers in Education*, *20*(3), 245–270. https://doi.org/10.1080/02671520500193744

Harlen, W. (2005b). Teachers' summative practices and assessment for learning - tensions and synergies. *The Curriculum Journal*, *16*(2), 207–223. https://doi.org/http://dx.doi.org/10.1080/09585170500136093

Harlen, W. (2006). On the relationship between assessment for formative and summative purposes. In W. Harlen & J. Gardner (Eds.), *Assessment and learning* (pp.87-102). London, England: Sage Publications.

Harlen, W. (2007). *Assessment of learning*. London, England: Sage Publications.

Harlen, W., & Deakin Crick, R. (2002). A systematic review of the impact of summative assessment and test on students' motivation for learning. In *Research evidence in Education Library, Issue 1*. London, England: University of London.

Harlen, W., & Deakin Crick, R. (2003). Testing and motivation for learning. *Assessment in Education: Principles, policy* & practice, *10*(2), 169–207. http://dx.doi.org/10.1080/0969594032000121270

Harlen, W., & James, M. (1996). Creating a positive impact of assessment on learning. Paper presented at the AERA annual conference, New York, NY. Retrieved from http://files.eric.ed.gov/fulltext/ED397137.pdf

Harris, L. R., & Brown, G. T. L. (2009). The complexity of teachers' conceptions of assessment: Tensions between the needs of schools and students. *Assessment in Education: Principles, Policy & Practice*, *16*(3), 365–381. https://doi.org/10.1080/09695940903319745

Harteis, C., Gruber, H., & Lehner, F. (2006). Epistemological beliefs and their impact on work, subjectivity and learning. In S. Billett, T. J. Fenwick, & M. Somerville (Eds.), *Work, subjectivity and learning: Understanding learning through working life* (pp. 123–140). Dordrecht, The Netherlands: Springer.

Hashweh, M. Z. (1987). Effects of subject-matter knowledge in the teaching of biology and physics. *Teaching and Teacher Education*, *3*(2), 109–120. https://doi.org/10.1016/0742-051X(87)90012-6

Hattie, J., & Peddie, R. (2003). School reports: 'Praising with faint damns'. *Set: Research Information for Teachers*, *3*, 4–9.

Hay, P., & Penney, D. (2012). *Assessment in physical education: A socio-cultural perspective*. Florence, KY: Taylor and Francis.

Heron, M., Hucker, J., Rooney, C., Robinson, V., Lai, M. K., & Mose, K. (2001). Progressive achievement testing at Southern Cross Middle School. *Set: Research Information for Teachers*, *2*, 34–38.

Hickey, D. T., & Anderson, K. T. (2007). Situative approaches to student assessment: Contextualizing evidence to transform practice. *Yearbook of the National Society for the Study of Education*, *106*(1), 264–287. https://doi.org/http://dx.doi.org/10.1111/j.1744-7984.2007.00105.x

Hill, D., & Kumar, R. (2012). *Global neoliberalism and education and its consequences*. New York, NY: Routledge.

Hill, M. (2000). Dot, slash, cross: How assessment can drive teachers to ticking instead of teaching. *Set: Research Information for Teachers*, *1*, 21–25.

Hill, A. (2006, September). Comparing apples with pears. Teacher knowledge about assessment and its impact on their practices. Paper presented at the British Educational Research Association Annual Conference, University of Warwick, England.

Hill, M., Cowie, B., Gilmore, A., & Smith, L. (2010). Preparing assessment-capable teachers: What should pre-service teachers know and be able to do? *Assessment Matters*, *2*, 43–64.

Hill, M., Grudnoff, L., & Ell, F. (2011). Learning to think as an effective mathematics teacher: Teacher educator impacts on curriculum knowledge and learning to teach. *Teachers and Curriculum*, *12*, 13–21. https://doi.org/http://dx.doi.org/10.15663/tandc.v12i1.26

Hill, M., Smith, L., Cowie, B., Gilmore, A., & Gunn, A. (2013). *Preparing initial primary and early childhood teacher education students to use assessment.* (Final summary report). Wellington, New Zealand: TLRI.

Hipkins, R. (2010). *The evolving NCEA*. Wellington, New Zealand: New Zealand Council for Educational Research.

Hipkins, R. (2012). Assessment of naturally occurring evidence of literacy. *Assessment Matters*, *4*, 95–109.

Hipkins, R. (2013). *NCEA one decade on*. Wellington, New Zealand: New Zealand Council for Educational Research.

Hipkins, R., Barker, M., & Bolstad, R. (2005). Teaching the 'nature of science': modest adaptations or radical reconceptions? *International Journal of Science Education*, *27*(2), 243–254. https://doi.org/10.1080/0950069042000276758

Hipkins, R., & Neill, A. (2006). *Shifting balances: The impact of Level 1 NCEA on the teaching of mathematics and science*. Wellington, New Zealand: Ministry of Education.

Hipkins, R., & Robertson, S. (2011). *Moderation and teacher learning: What can research tell us about their interrelationships?* Wellington, New Zealand: New Zealand Council for Educational Research.

Hipkins, R., Vaughan, K., Beals, F., Ferral, H., & Gardiner, B. (2005). *Shaping our futures: Meeting secondary students' learning needs in a time of evolving qualifications*. Wellington, New Zealand: New Zealand Council for Educational Research.

Ho, H.-Z., O'Farrell, S. L., Hong, S., & You, S. (2006). Developmental research: Theory, method, design and statistical analysis. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 207–225). Washington, DC: Lawrence Erlbaum Associates.

Hodge, L. L., & Cobb, P. (2016). Two views of culture and their implications for mathematics teaching and learning. *Urban Education*, 1–25. https://doi.org/10.1177/0042085916641173

Hoekstra, A., Beijaard, D., Brekelmans, M., & Korthagen, F. (2007). Experienced teachers' informal learning from classroom teaching. *Teachers and Teaching*, *13*(2), 191–208. https://doi.org/10.1080/13540600601152546

Hoffman, J. V., Roller, C., Maloch, B., Sailors, M., Duffy, G., & Beretvas, S. N. (2005). Teachers' preparation to teach reading and their experiences and practices in the first three years of teaching. *The Elementary School Journal*, *105*(3), 267–287. https://doi.org/http://dx.doi.org/10.1086/428744

Hollingsworth, S. (1989). Prior beliefs and cognitive change in learning to teach. *American Educational Research Journal*, *26*(2), 160–189. https://doi.org/10.3102/00028312026002160

Hope, J., & Townsend, M. (1983). Student teachers' understanding of science concepts. *Research in Science Education*, *13*, 177–183. https://doi.org/10.1007/BF02356704

Horn, I. S., Nolen, S. B., Ward, C., & Campbell, S. S. (2008). Developing practices in multiple worlds: The role of identity in learning to teach. *Teacher Education Quarterly*, *35*(3), 61–72.

Hornby, G., & Witte, C. (2014). Ability grouping in New Zealand high schools: Are practices evidence-based? *Preventing School Failure*, *58*(2), 90–95. https://doi.org/10.1080/1045988X.2013.782531

Hornby, G., Witte, C., & Mitchell, D. (2011). Policies and practices of ability grouping in New Zealand intermediate schools. *Support for Learning*, *26*(3), 92–96. https://doi.org/10.1111/j.1467-9604.2011.01485.x

Horwitz, E. K. (1985). Using student beliefs about language learning and teaching in the foreign language methods course. *Foreign Language Annals*, *18*(4), 333–340. https://doi.org/10.1111/j.1944-9720.1985.tb01811.x

Houseal, A. K., Abd-El-Khalick, F., & Destefano, L. (2014). Impact of a student-teacher-scientist partnership on students' and teachers' content knowledge, attitudes toward science, and pedagogical practices. *Journal of Research in Science Teaching*, *51*(1), 84–115. https://doi.org/10.1002/tea.21126

Huberman, M. (1991). Survivre à la première phase de la carrière. *Cahiers Pedagogiques*, *290*, 15–17.

Hume, A. (2006). Carrying out a practical investigation: The student-experienced curriculum (Doctoral thesis, University of Waikato, Hamilton, New Zealand). Retrieved from http://researchcommons.waikato.ac.nz/handle/10289/3991

Hume, A. (2010). CoRes and PaP-eRs. *New Zealand Science Teacher*, *124*, 38–40.

Hume, A., & Berry, A. (2010). Constructing CoRes - A strategy for building PCK in pre-service science teacher education. *Research in Science Education*, *41*(3), 341–355. https://doi.org/10.1007/s11165-010-9168-3

Hume, A., & Coll, R. (2008). Student experiences of carrying out a practical science investigation under direction. *International Journal of Science Education*, *30*(9), 1201–1228. https://doi.org/10.1080/09500690701445052

Hume, A., & Coll, R. (2009). Assessment of learning, for learning, and as learning: New Zealand case studies. *Assessment in Education: Principles, Policy & Practice*, *16*(3), 269–290. https://doi.org/10.1080/09695940903319661

Hume, A., & Coll, R. (2010). Authentic student inquiry: The mismatch between the intended curriculum and the student-experienced curriculum. *Research in Science & Technological Education*, *28*(1), 43–62. https://doi.org/10.1080/02635140903513565

Hursh, D. (2007). Assessing No Child Left Behind and the rise of neoliberal education policies. *American Educational Research Journal*, *44*(3), 493–518. https://doi.org/10.3102/0002831207306764

Hutchings, M. (2015). Exam factories? The impact of accountability measures on children and young people. London, England: National Union of Teachers.

Hutner, T. L., & Markman, A. B. (2016). Proposing an operational definition of science teacher beliefs. *Journal of Science Teacher Education*, *27*, 675–691. https://doi.org/10.1007/s10972-016-9480-5

Immordino-Yang, M. H. (2011). Implications of affective and social neuroscience for educational theory. *Educational Philosophy and Theory*, *43*(1), 98–103. https://doi.org/10.1111/j.1469-5812.2010.00713.x

Immordino-Yang, M. H., & Damasio, A. (2007). We feel, therefore we learn: The relevance of affective and social neuroscience to education. *Mind, Brain, and Education*, *1*(1), 3–10. https://doi.org/10.1111/j.1751-228X.2007.00004.x

Ireson, J., Clark, H., & Hallam, S. (2002). Constructing ability groups in the secondary school: Issues in practice. *School Leadership & Management*, *22*(2), 163–176. https://doi.org/10.1080/1363243022000007737

Irving, S. E., Harris, L. R., & Peterson, E. R. (2011). 'One assessment doesn't serve all the purposes' or does it? New Zealand teachers describe assessment and feedback. *Asia Pacific Education Review*, *12*(3), 413–426. https://doi.org/http://dx.doi.org.ezproxy.waikato.ac.nz/10.1007/s12564-011-9145-1

Isenbarger, L., & Zembylas, M. (2006). The emotional labour of caring in teaching. *Teaching and Teacher Education*, *22*, 120–134.

Izci, K. (2013). *Investigating high school chemistry teachers' perceptions, knowledge and practice of classroom assessment* (Doctoral dissertation, University of Missouri). Retrieved from https://mospace.umsystem.edu/xmlui/handle/10355/44057

Jacobs, J., Gregory, A., Hoppey, D., & Yendol-Hoppey, D. (2009). Data literacy: Understanding teachers data use in the context of accountability and response to intervention. *Action in Teacher Education*, *31*(3), 41–55. https://doi.org/10.1080/01626620.2009.10463527

James, M. (1998). *Using assessment for school improvement*. Oxford, England: Heinemann Educational.

James, M., & Pedder, D. (2006). Professional learning as a condition for assessment for learning. In J. Gardner (Ed.), *Assessment and learning*. London, England: Sage Publications.

Jennings, J. (2008). *A composer-teacher in context: Music for the performing arts faculty in a New Zealand secondary school* (Doctoral thesis, University of Waikato, Hamilton, New Zealand). Retrieved from http://researchcommons.waikato.ac.nz/handle/10289/2605

Jensen, S., McKinley, E., & Madyar, I. (2010). NCEA subject choices in mid-low-decile schools: 'What schools and parents need to know about the university pathway'. *Set: Research Information for Teachers*, *2*, 37–43.

Johnson, S. (2011). *Assessing learning in the primary classroom*. London, England: Routledge.

Johnson, S. (2013). On the reliability of high-stakes teacher assessment. *Research Papers in Education*, *28*(1), 91–105. https://doi.org/10.1080/02671522.2012.754229

Johnson, W., Nyamekye, F., Chazan, D., & Rosenthal, B. (2013). Teaching with speeches: A black teacher who uses the mathematics classroom to prepare students for life. *Teachers College Record*, *115*(2), 1–26.

Jones, A., & Buntting, C. (2013). International, national and classroom assessment: Potent factors in shaping what counts in school science. In D. Corrigan, R. Gunstone, & A. Jones (Eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy* (pp. 33–53). Dordrecht, The Netherlands: Springer Netherlands.

Jones, A., & Moreland, J. (2005). The importance of pedagogical content knowledge in assessment for learning practices: A case-study of a whole-school approach. *The Curriculum Journal*, *16*(2), 193–206. https://doi.org/10.1080/09585170500136044

Jones, E. (2010). Personal theory and reflection in a professional practice portfolio. *Assessment and Evaluation in Higher Education*, *35*(6), 699–710. https://doi.org/http://dx.doi.org/10.1080/02602930902977731

Kagan, D. M. (1992). Professional growth among preservice and beginning teachers. *Review of Educational Research*, *62*(2), 129–169. https://doi.org/10.2307/1170578

Kahl, S., Hofman, P., & Bryant, S. (2013). *Assessment literacy standards and performance measures for teacher candidates and practising teachers* (pp. 1–31). Dover, NH: Measured progress.

Kang, E. J. S., Bianchini, J. A., & Kelly, G. J. (2012). Crossing the border from science student to science teacher: Preservice teachers' views and experiences learning to teach inquiry. *Journal of Science Teacher Education*, *24*(3), 427–447. https://doi.org/10.1007/s10972-012-9317-9

Kayumova, S., & Tippins, D. (2016). Toward re-thinking science education in terms of affective practices: Reflections from the field. *Cultural Studies of Science Education*, *11*(3), 567–575. https://doi.org/10.1007/s11422-015-9695-3

Keddie, A. (2016). Children of the market: performativity, neoliberal responsibilisation and the construction of student identities. *Oxford Review of Education*, *42*(1), 108–122. https://doi.org/10.1080/03054985.2016.1142865

Kelchtermans, G. (1996). Teacher vulnerability: Understanding its moral and political roots. *Cambridge Journal of Education*, *26*(3), 307–323. https://doi.org/10.1080/0305764960260302

Kelchtermans, G. (2005). Teachers' emotions in educational reforms: Self-understanding, vulnerable commitment and micropolitical literacy. *Teaching and Teacher Education*, *21*(8), 995–1006. https://doi.org/10.1016/j.tate.2005.06.009

Kelly, P. (2006). What is teacher learning? A socio-cultual perspective. *Oxford Review of Education*, *32*(4), 505–519. https://doi.org/http://dx.doi.org/10.1080/03054980600884227

Kennedy, M. M. (1999). The role of preservice teacher education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (1st ed.). San Francisco, CA: Jossey-Bass.

Kieran, T. E. (2000). Dichotomies or binoculars: Reflections on the papers by Steffe and Thompson and by Lerman. *Journal for Research in Mathematics Education*, *31*, 228–233. https://doi.org/10.2307/749753

Kind, V. (2009). Pedagogical content knowledge in science education: Perspectives and potential for progress. *Studies in Science Education*, *45*(2), 169–204. https://doi.org/10.1080/03057260903142285

Kind, V. (2014). Science teachers' content knowledge. In H. Venkat, M. Rollnick, J. Loughran, & M. Askew (Eds.), *Exploring mathematics and science teachers' knowledge: Windows into teacher thinking* (pp. 15–28). New York, NY: Routledge.

Kind, V. (2015). On the beauty of knowing then not knowing: Pinning down the illusive qualities of PCK. In Amanda Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 178–196). New York, NY: Routledge.

Klenowski, V., & Wyatt-Smith, C. (2014). *Assessment for education*. London, England: Sage Publications.

Kofoed, W. J. (2009). *Written reporting: Strengthening learning partnerships through purposeful reporting* (Unpublished EdD thesis, University of Auckland, New Zealand).

Koh, K. H. (2011). Improving teachers' assessment literacy through professional development. *Teaching Education*, *22*(3), 255–276. https://doi.org/10.1080/10476210.2011.593164

Koh, K. H. (2014). Authentic assessment, teacher judgment and moderation in a context of high accountability. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning* (pp. 249–264). Dordrecht, The Netherlands: Springer.

Korthagen, F. A. J. (2010). Situated learning theory and the pedagogy of teacher education: Towards an integrative view of teacher behavior and teacher learning. *Teaching and Teacher Education*, *26*(1), 98–106. https://doi.org/10.1016/j.tate.2009.05.001

Kozma, R. (2009). Transforming education: Assessing and teaching 21st century skills. In E. Scheuermann & J. Bjjörnsson (Eds.), *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing* (pp. 13–23). Luxembourg: European Commission Joint Research Centre.

Kruse, R. A., & Roehrig, G. H. (2005). A comparison study: Assessing teachers' conceptions with the chemistry concepts inventory. *Journal of Chemical Education*, *82*(8), 1246. https://doi.org/10.1021/ed082p1246

Kvale, S. (2007). *Doing interviews*. Thousand Oaks, CA: Sage Publications.

Lai, K., Cabrera, J., Vitale, J. M., Madhok, J., Tinker, R., & Linn, M. C. (2016). Measuring graph comprehension, critique, and construction in science. *Journal of Science Education and Technology*, *25*(4), 665–681. https://doi.org/10.1007/s10956-016-9621-9

Lai, M. K., & McNaughton, S. (2013). Analysis and discussion of classroom and achievement data to raise student achievement. In K. Schildkamp, M. K. Lai, & L. Earl (Eds.), *Data-based decision making in education:*

Challenges and opportunities (Vol. 17, pp. 23–47). Dordrecht, The Netherlands: Springer.

Lane, S. (2013). Performance assessment. In J.H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 313–330). Los Angeles, CA: Sage Publications.

Langdon, F. J. (2007). *Beginning teacher learning and professional development: An analysis of induction programmes* (Doctoral thesis, University of Waikato, Hamilton, New Zealand). Retrieved from http://hdl.handle.net/10289/2624

Lasky, S. (2005). A sociocultural approach to understanding teacher identity, agency and professional vulnerability in a context of secondary school reform. *Teaching and Teacher Education*, *21*(8), 899–916. https://doi.org/10.1016/j.tate.2005.06.003

Lather, P. (1991). *Getting smart: Feminist research and pedagogy with/in the postmodern*. New York, NY: Routledge.

Lave, J. (1996). Teaching, as learning, in practice. *Mind, Culture, and Activity*, *3*(3), 149–164. https://doi.org/10.1207/s15327884mca0303_2

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, England: Cambridge University Press.

Leach, J., & Scott, P. (2003). Individual and sociocultural views of learning in science education. *Science and Science Education*, *12*, 91–113.

Lederman, N. G. (2007). Nature of science: Past, present, and future. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 831–880). London, England: Lawrence Erlbaum Associates.

Leighton, J. P., Gokiert, R. J., Cor, M. K., & Heffernan, C. (2010). Teacher beliefs about the cognitive diagnostic information of classroom- versus large-scale tests: implications for assessment literacy. *Assessment in Education: Principles, Policy & Practice*, *17*(1), 7–21. https://doi.org/10.1080/09695940903565362

Levin, B. (2014). The development of teachers' beliefs. In H. Fives & M. G. Gill (Eds.), *International handbook of research on teachers' beliefs* (pp. 48–65). New York, NY: Routledge.

Levin, B., & He, Y. (2008). Investigating the content and sources of teacher candidates' personal practical theories (PPTS). *Journal of Teacher Education*, *59*(1), 55–68. https://doi.org/10.1177/0022487107310749

Lin, T.-J., Lee, M.-H., & Tsai, C.-C. (2014). The commonalities and dissonances between high-school students' and their science teachers' conceptions of science learning and conceptions of science assessment: A Taiwanese sample study. *International Journal of Science Education*, *36*(3), 382–405. https://doi.org/10.1080/09500693.2013.780317

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverley Hills, CA: Sage Publications.

Lincoln, Y. S., Lynham, S. A., & Guba, E. G. (2011). Paradigmatic controversies, contradictions, and emerging confluences, revisited. In N. Denzin & Y. Lincoln (Eds.), *The Sage handbook of qualitative research* (4th ed.). Thousand Oaks, CA: Sage Publications.

Locke, T. (2005). Talking across the divide: English teachers respond to NCEA. *Waikato Journal of Education*, *11*(2), 113–136. https://doi.org/10.15663/wje.v11i2.313

Lomax, R. G. (1996). On becoming assessment literate: An initial look at preservice teachers' beliefs and practices. *The Teacher Educator*, *31*(4), 292–303. https://doi.org/10.1080/08878739609555122

Longworth, E., & McBride, T. (1994). *The privacy act: A guide*. Wellington, New Zealand: GP Publication.

Looney, J. W. (2011). *Integrating formative and summative assessment: Progress towards a seamless system?* (OECD Education Working Papers No. 58). Paris, France: OCED Publishing. http://dx.doi.org/10.1787/5kghx3kbl734-en

Lortie, D. (1975). *Schoolteacher: Sociological study*. Chicago, IL: University of Chicago.

Loughran, J. (2006). *Developing a pedagogy of teacher education: Understanding teaching and learning about teaching*. London, England: Routledge.

Loughran. J. (2013). *Developing a pedagogy of teacher education: Understanding teaching and learning about teaching*. London, England: Routledge.

Loughran, J., Berry, A., & Mulhall, P. (2006). *Understanding and developing science teachers' pedagogical content knowledge*. Rotterdam, The Netherlands: Sense Publishers.

Loughran, J., Mulhall, P., & Berry, A. (2008). Exploring pedagogical content knowledge in science teacher education. *International Journal of Science Education*, *30*(10), 1301–1320. https://doi.org/http://dx.doi.org/10.1080/09500690802187009

Lovett, S., & Sinclair, L. (2005). *The socialisation of teachers into the culture of assessment*. Wellington, New Zealand: New Zealand Council for Educational Research.

Luft, J. A. (2009). Beginning secondary science teachers in different induction programmes: The first year of teaching. *International Journal of Science Education*, *31*(17), 2355–2384. https://doi.org/10.1080/09500690802369367

Luft, J. A., Dubois, S. L., Nixon, R. S., & Campbell, B. K. (2015). Supporting newly hired teachers of science: Attaining teacher professional standards. *Studies in Science Education*, *51*(1), 1–48. https://doi.org/10.1080/03057267.2014.980559

Luft, J. A., Hill, K., Weeks, C., Raven, S., & Nixon, R. S. (2013, April). The knowledge needed for teaching science: A study of in and out-of-field science teachers. Paper presented at the National Association for Research

in Science Teaching, Rio Grande Puerto Rico. Retrieved from
http://www.researchgate.net/profile/Ryan_Nixon/publication/259780270_
Science_Knowledge_for_Teaching_Characterizations_from_Early_Career
_Teachers/file/6a85e52dd40f721de7.pdf

Lukin, L. E., Bandalos, D. L., Eckhout, T. J., & Mickelson, K. (2004). Facilitating
the development of assessment literacy. *Educational Measurement: Issues
and Practice*, *23*(2), 26–32. https://doi.org/10.1111/j.1745-
3992.2004.tb00156.x

Lyon, E. G. (2011). Beliefs, practices, and reflection: Exploring a science
teacher's classroom assessment through the assessment triangle model.
Journal of Science Teacher Education, *22*, 417–435.
https://doi.org/10.1007/s10972-011-9241-4

Lyon, E. G. (2012). *Unravelling the complex: Changes in secondary science
preservice teachers' assessment expertise* (Doctoral dissertation,
University of California). Retrieved from
http://escholarship.org/uc/item/65w699dc

Lyon, E. G. (2013a). Conceptualizing and exemplifying science teachers'
assessment expertise. *International Journal of Science Education*, *35*(7),
1208–1229. https://doi.org/10.1080/09500693.2013.770180

Lyon, E. G. (2013b). Learning to assess science in linguistically diverse
classrooms: Tracking growth in secondary science preservice teachers'
assessment expertise. *Science Education*, *97*(3), 442–467.
https://doi.org/10.1002/sce.21059

Lyon, E. G. (2013c). What about language while equitably assessing science?:
Case studies of preservice teachers' evolving expertise. *Teaching and
Teacher Education*, *32*, 1–11. https://doi.org/10.1016/j.tate.2012.12.006

Ma, H. (2013). Science teachers' understanding of the college entrance
examination in a climate of national reform in China. In D. Corrigan, R.
Gunstone, & A. Jones (Eds.), *Valuing assessment in science education:
Pedagogy, curriculum, policy*. Dordrecht, The Netherlands: Springer.

Maclellan, E. (2004). Initial knowledge states about assessment: Novice teachers'
conceptualisations. *Teaching and Teacher Education*, *20*(5), 523–535.
https://doi.org/10.1016/j.tate.2004.04.008

Madyar, I., McKinley, E., Jensen, S., & van der Merwe, A. (2009). *Towards
university: Navigating NCEA course choices in low-mid decile schools*.
Auckland, New Zealand: University of Auckland. Retrieved from
http://hdl.handle.net/2292/16707

Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources and development
of pedagogical content knowledge for science teaching. In J. Gess-
Newsome & N. G. Lederman (Eds.), *Examining pedagogical content
knowledge* (pp. 95–132). Dordrecht, The Netherlands: Kluwer.

Mahuika, R., Berryman, M., & Bishop, R. (2011). Issues of culture and
assessment in New Zealand education pertaining to Māori students.
Assessment Matters 3, 183-198.

Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher*, *42*(1), 30–37. https://doi.org/10.3102/0013189X12459803

Mandinach, E. B., & Gummer, E. S. (2016). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education*, *60*, 366–376. https://doi.org/10.1016/j.tate.2016.07.011

Mansell, W., James, M., & Assessment Reform Group. (2009). *Assessment in schools: Fit for purpose? A commentary by the Teaching and Learning Research Programme*. London, England: Economic and Social Research Council.

Mansour, N., EL-Deghaidy, H., Alshamrani, S., & Aldahmash, A. (2014). Rethinking the theory and practice of continuing professional development: Science teachers' perspectives. *Research in Science Education*, *44*, 949–973. https://doi.org/http://dx.doi.org/10.1007/s11165-014-9409-y

Maxwell, G. (2007). *Implications for moderation of proposed changes to senior secondary school syllabuses* (Paper commissioned by the Queensland Studies Authority). Brisbane, Australia: Queensland Studies Authority.

McCaslin, M., & Legg Burross, H. L. (2011). Research on individual differences within a sociocultural perspective: Co-regulation and adaptive learning. *Teachers College Record*, *113*(2), 325–349.

McDonald, S., Warren, E., & DeVries, E. (2011). Refocusing on oral language and rich representations to develop the early mathematical understandings of indigenous students. *The Australian Journal of Indigenous Education*, *40*, 9–17. https://doi.org/10.1375/ajie.40.9

McGee, J., & Colby, S. (2014). Impact of an assessment course on teacher candidates' assessment literacy. *Action in Teacher Education*, *36*(5–6), 522–532. https://doi.org/10.1080/01626620.2014.977753

McGinnis, J. R., Parker, C., & Graeber, A. O. (2004). A cultural perspective of the induction of five reform-minded beginning mathematics and science teachers. *Journal of Research in Science Teaching*, *41*, 729–747. https://doi.org/10.1002/tea.20022

McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualisation. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 575–596). Reston, VA: National Council of Teachers of Mathematics.

McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, *20*(1), 20–32.

McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, *22*(4), 34–43. https://doi.org/10.1111/j.1745-3992.2003.tb00142.x

McMillan, J. H. (2005). *The impact of high-stakes test results on teachers' instructional and classroom practices.* Metropolitan Educational Research Consortium, Virginia Commonwealth University, Richmond, VA. ERIC Document Reproduction Service No. ED490648.

McMillan, J. H. (2013). Why we need research on classroom assessment. In J.H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 3–16). Los Angeles, CA: Sage Publications.

McMillan, J. H., & Nash, S. (2000, April). Teacher classroom assessment and grading practices decision making. Paper presented at the Annual meeting of National Council on Measurement in Education, New Orleans, LA.

Meis Friedrichsen, P., & Dana, T. M. (2005). Substantive-level theory of highly regarded secondary biology teachers' science teaching orientations. *Journal of Research in Science Teaching*, *42*(2), 218–244. https://doi.org/10.1002/tea.20046

Merriam, S. B. (1988). *Case study research in education: A qualitative approach.* San Francisco, CA: Jossey-Bass.

Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco, CA: Jossey-Bass.

Mertler, C. A. (2005). Secondary Teachers' Assessment Literacy: Does Classroom Experience Make a Difference? *American Secondary Education*, *33*(2), 76–92.

Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools*, *12*(2), 101–113. https://doi.org/10.1177/1365480209105575

Mertler, C., & Campbell, C. (2005, April). Measuring teachers' knowledge and application of classroom assessment concepts: Development of the Assessment Literacy Inventory. Paper presented at the Annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.

Messick, S. (1994). The Interplay of Evidence and Consequences in the Validation of Performance Assessments. *Educational Researcher*, *23*(2), 13–23. https://doi.org/10.2307/1176219

Meyer, L. H., McClure, J., Walkey, F., Weir, K. F., McKenzie, L. (2009). Secondary student motivation orientations and standards-based achievement outcomes. *British Journal of Educational Psychology*, *79*(2), 273–293. https://doi.org/10.1348/000709908X354591

Meyer, L. H., Weir, K. F., McClure, J., Walkey, F., & McKenzie, L. (2007). *Longitudinal research on the relationship between NCEA and student motivation and achievement: Final report to the Ministry of Education.* Wellington, New Zealand: Ministry of Education and Victoria University.

Meyer, L. H., Weir, K. F., McClure, J., Walkey, F., & McKenzie, L. (2009). *Longitudinal research on the NCEA and student motivation and achievement: A three-year follow-up. Final report to Ministry of*

Education. Wellington, New Zealand: Ministry of Education and Victoria University.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Miles, M. B., Huberman, A. M., & Saldana, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Millar, R. (2013). Improving science education: Why assessment matters. In D. Corrigan, R. Gunstone, & A. Jones (Eds.), *Valuing assessment in science education: Pedagogy, curriculum, policy* (pp. 55–68). Dordrecht, The Netherlands: Springer.

Figure 5.   Ministry of Education. (2007). The New Zealand curriculum for English-medium teaching and learning in years 1-13. Wellington, New Zealand: Learning Media.

Figure 6.   Ministry of Education. (2008). Te Marautanga o Aotearoa. Wellington, New Zealand: Learning Media.

Ministry of Education. (2009). Relevant legislation, NEGs and NAGs Retrieved from http://www.minedu.govt.nz

Ministry of Education. (2011a). Assessment: Schooling sector. Wellington, New Zealand: Ministry of Education

Ministry of Education. (2011b). *OECD review on evaluation and assessment frameworks for improving school outcomes: Country background report for New Zealand.* Wellington, New Zealand: Ministry of Education.

Ministry of Education. (2015). *The National Education Goals.* Retrieved from https://www.education.govt.nz/ministry-of-education/legislation/negs/

Mislevy, R. J. (1993). Foundations of a new test theory. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests.* Hillsdale, NJ: Erlbaum.

Mizutani, S., Rubie-Davies, C., Hattie, J., & Philp, J. (2011). Do beliefs about NCEA and its washback effects vary depending on subject? *New Zealand Journal of Educational Studies*, *46*(2), 47–59.

Moeed, A. (2013). Science investigation that best supports student learning: Teachers' understanding of science investigation. *International Journal of Environmental and Science Education*, *8*(4), 537–559.

Moeed, A. (2015). *Science Investigation: Student Views about Learning, Motivation and Assessment.* Singapore: Springer.

Moeed, A., & Hall, C. (2011). Teaching, learning and assessment of science investigation in Year 11: Teachers' response to NCEA. *New Zealand Science Review*, *68*(3), 95–102.

Moreland, J., Jones, A., & Chambers, M. (2001). Enhancing student learning in technology through enhancing teacher formative interactions. *Set: Research Information for Teachers*, *3*, 16–19.

Morrison, K., & Tang, F. H. J. (2002). Testing to destruction: A problem in a small state. *Assessment in Education: Principles, Policy & Practice*, *9*(3), 289–317. https://doi.org/10.1080/0969594022000027654a

Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation, 7*(10), 1–10.

Moss, C. M. (2013). Research on classroom summative assessment. In J.H. McMillan (Ed.), *Sage handbook on research on classroom assessment* (pp. 235–255). Thousand Oaks, CA: Sage Publications.

Moss, P. A., Pullin, D. C., Gee, J. P., Haertel, E. H., & Young, L. J. (Eds.). (2008). *Assessment, equity and opportunity to learn*. Cambridge, England: Cambridge University Press.

Mueller, A., & Skamp, K. (2003). Teacher candidates talk: Listen to the unsteady beat of learning to teach. *Journal of Teacher Education*, *54*(5), 428–440. https://doi.org/10.1177/0022487103256902

Murray, M. (2013). *Associate teachers as successful mentors; associate teachers' and student teachers' perspectives* (Master of Teaching and Learning thesis, University of Canterbury, Christchurch, New Zealand). Retrieved from http://ir.canterbury.ac.nz/bitstream/10092/7898/1/thesis_fulltext.pdf

Näring, G., Briët, M., & Brouwers, A. (2006). Beyond demand–control: Emotional labour and symptoms of burnout in teachers. *Work & Stress*, *20*(4), 303–315. https://doi.org/10.1080/02678370601065182

National Curriculum Task Group on Assessment and Testing. (1988). *A report*. London, United Kingdom: HMSO.

National Research Council. (2010). *Preparing teachers: Building evidence for sound policy*. Washington, DC: National Academies Press.

Neville, B. (2013). The enchanted loom. In M. Newberry, A. Gallant, & P. Riley (Eds.), *Emotion and school: Understanding how the hidden curriculum influences relationship leadership teaching and learning*. Bingley, England: Emerald.

New Zealand Qualifications Authority. (2001). Learning and assessment: A guide to assessment for the national qualifications framework. New Zealand Qualifications Authority. Retrieved from http://www.nzqa.govt.nz/assets/Providers-and-partners/Assessment-and-moderation/Assessment-of-standards/learningassessment.pdf

New Zealand Qualifications Authority. (2015a). Internal moderation. Retrieved from http://www.nzqa.govt.nz/providers-partners/assessment-and-moderation/managing-national-assessment-in-schools/secondary-moderation/external-moderation/internal-moderation/

New Zealand Qualifications Authority. (2015b). *NZQA rules*. Retrieved from: www.nzqa.govt.nz.

New Zealand Qualifications Authority. (2016a). Special assessment conditions. Retrieved from http://www.nzqa.govt.nz/providers-partners/assessment-

and-moderation/managing-national-assessment-in-schools/special-assessment-conditions/

New Zealand Qualifications Authority. (2016b). *Statement of Intent 2016/17-2019/20*. Wellington, New Zealand: New Zealand Qualifications Authority.

New Zealand Qualifications Authority. (2017). Secondary school statistics. Retrieved from http://www.nzqa.govt.nz/studying-in-new-zealand/secondary-school-and-ncea/find-information-about-a-school/secondary-school-statistics/

Newby, P. (2014). *Research methods for education*. London, England: Routledge.

Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, *14*(2), 149–170. https://doi.org/10.1080/09695940701478321

Nias, J. (1989). *Primary teachers talking: A study of teaching as work*. London, England: Routledge.

Nichols, S. L., & Harris, L. R. (2016). Accountability assessment's effects on teachers and schools. In G. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 40–56). New York, NY: Routledge.

Nolen, S. B., Horn, I. S., Ward, C. J., & Childers, S. A. (2011). Novice teacher learning and motivation across contexts: Assessment tools as boundary objects. *Cognition and Instruction*, *29*(1), 88–122. https://doi.org/10.1080/07370008.2010.533221

Nolen, S. B., Ward, C. J., & Horn, I. S. (2011). Methods for taking a situative approach to studying the development of motivation, identity, and learning in multiple social contexts. *European Journal of Psychology of Education*, *27*(2), 267–284. https://doi.org/10.1007/s10212-011-0086-1

Nusche, D., Laveault, D., MacBeath, J., & Santiago, P. (2012). OECD reviews of evaluation and assessment in education: New Zealand, main conclusions. Retrieved from http://www.oecd.org/education/school/49681563.pdf

O'Connor, K. E. (2008). 'You choose to care': Teachers, emotions and professional identity. *Teaching and Teacher Education*, *24*, 117–126. https://doi.org/10.1016/j.tate.2006.11.008

Ogan-Bekiroglu, F., & Suzuk, E. (2014). Pre-service teachers' assessment literacy and its implementation into practice. *Curriculum Journal*, *25*(3), 344–371. https://doi.org/10.1080/09585176.2014.899916

O'Neill, O. (2013). Intelligent accountability in education. *Oxford Review of Education*, *39*(1), 4–16. https://doi.org/10.1080/03054985.2013.764761

Oppenheim, A. N. (1992). *Questionnaire design, interviewing and attitude measurement*. London, England: Pinter Publishers

Organisation for Economic Co-operation and Development. (2013). *Synergies for better learning: An international perspective on evaluation and*

assessment (OECD reviews of evaluation and assessment). Paris, France: OECD Publishing.

Otrel-Cass, K., Cowie, B., & Glynn, T. (2009). Connecting science teachers with their Māori students. *Set: Research Information for Teachers*, *2*, 34–41.

Ovens, A. (1996). Field experience: An analysis of students' problems. *Journal of Physical Education New Zealand*, *29*(1), 17–22.

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, *62*(3), 307–332. https://doi.org/10.3102/00346543062003307

Panizzon, D., & Pegg, J. (2008). Assessment practices: Empowering mathematics and science teachers in rural secondary schools to enhance student learning. *International Journal of Science & Mathematics Education*, *6*(2), 417–436. https://doi.org/10.1007/s10763-007-9084-9

Park, S., & Chen, Y.-C. (2012). Mapping out the integration of the components of pedagogical content knowledge (PCK): Examples from high school biology classrooms. *Journal of Research in Science Teaching*, *49*(7), 922–941. https://doi.org/10.1002/tea.21022

Park, S., & Oliver, J. S. (2008). Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education*, *38*(3), 261–284. https://doi.org/10.1007/s11165-007-9049-6

Patton, M. Q. (1990). *Quality evaluation and research methods*. Newbury Park, CA: Sage Publications.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Patton, M. Q. (2015). *Qualitative research & evaluation methods : Integrating theory and practice* (4th ed.). Thousand Oaks, CA: Sage Publications.

Pedder, D., & James, M. (2012). Professional learning as a condition for Assessment for Learning. In J. Gardner (Ed.), *Assessment and learning* (2nd ed., pp. 33–48). London, England: Sage Publications.

Pekrun, R., Elliot, A. J., & Maier, M. A. (2006). Achievement goals and discrete achievement emotions: A theoretical model and prospective test. *Journal of Educational Psychology*, *98*(3), 583. https://doi.org/10.1037/0022-0663.98.3.583

Pekrun, R., Goetz, T., Perry, R. P., Kramer, K., Hochstadt, M., & Molfenter, S. (2004). Beyond test anxiety: Development and validation of the test emotions questionnaire (TEQ). *Anxiety, Stress, & Coping*, *17*(3), 287–316. https://doi.org/10.1080/10615800412331303847

Pekrun, R., & Linnenbrink-Garcia, L. (2014). *International handbook of emotions in education*. New York, NY: Routledge.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academies Press.

Penuel, W. R., & Shepard, L. A. (2017). Social models of learning and assessment. In A. A. Rupp & J. P. Leighton (Eds.), *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications* (pp. 146–173). Chichester, England: Wiley Blackwell.

Peressini, D., Borko, H., Romagnano, L., Knuth, E., & Willis, C. (2004). A conceptual framework for learning to teach secondary mathematics: A situative perspective. *Educational Studies in Mathematics*, *56*(1), 67–96. https://doi.org/10.1023/B:EDUC.0000028398.80108.87

Philips, D. (2003). Lessons from New Zealand's National Qualifications Framework. *Journal of Education and Work*, *16*(3), 289–304. https://doi.org/10.1080/1363908032000099458

Phoenix, C., & Sparkes, A. C. (2009). Being Fred: Big stories, small stories and the accomplishment of a positive ageing identity. *Qualitative Research*, *9*(2), 219–236. https://doi.org/10.1177/1468794108099322

Pierce, R., & Chick, H. (2010). Interpreting literacy and numeracy testing reports: What do teachers need to know? In C. Reading (Ed.), Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia. Voorburg, The Netherlands: International Statistical Institute.

Pigge, F. L., & Marso, R. N. (1997). A seven year longitudinal multi-factor assessment of teaching concerns and development through preparation and early years of teaching. *Teaching and Teacher Education*, *13*(2), 225–235. https://doi.org/http://dx.doi.org/10.1016/s0742-051x(96)00014-5

Plake, B., & Impara, J. (1997). Teacher assessment literacy: What do teachers know about assessment? In G. D. Phye (Ed.), *Handbook of classroom assessment* (pp. 55–68). New York, NY: Academic Press.

Pong, W. Y., & Chow, J. C. S. (2002). On the pedagogy of examinations in Hong Kong. *Teaching and Teacher Education*, *18*(2), 139–149. https://doi.org/10.1016/S0742-051X(01)00059-2

Pope, N., Green, S. K., Johnson, R. L., & Mitchell, M. (2009). Examining teacher ethical dilemmas in classroom assessment. *Teaching and Teacher Education*, *25*(5), 778–782. https://doi.org/10.1016/j.tate.2008.11.013

Popham, W. J. (2000). *Modern educational measurement: practical guidelines for educational leaders* (3rd ed.). Boston, MA: Allyn & Bacon.

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, *48*(1), 4–11. https://doi.org/10.1080/00405840802577536

Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, *46*(4), 265–273. https://doi.org/10.1080/08878730.2011.605048

Poskitt, J., & Taylor, K. (2008). *National education findings of Assess to Learn (AToL) report*. Wellington, New Zealand: Ministry of Education.

Postlethwaite, K., & Haggarty, L. (2012). Student teachers' thinking about learning to teach: A study of student teachers of mathematics and science at the end of their initial training. *Research Papers in Education*, *27*(3), 263–284. https://doi.org/http://dx.doi.org/10.1080/02671522.2010.501906

Powietrzyńska, M., & Gangji, A.-K. H. (2016). 'I understand why people need to ease their emotions': Exploring mindfulness and emotions in a conceptual physics classroom of an elementary teacher education program. *Cultural Studies of Science Education*, *11*(3), 693–712. https://doi.org/10.1007/s11422-016-9772-2

Pratt, N. (2016). Playing the levelling field: Teachers' management of assessment in English primary schools. *Assessment in Education: Principles, Policy & Practice*, 1–15. https://doi.org/10.1080/0969594X.2016.1264924

Pryor, J., & Crossouard, B. (2008). A socio-cultural theorisation of formative assessment. *Oxford Review of Education*, *34*(1), 1–20. https://doi.org/10.1080/03054980701476386

Pugach, M. C. (1992). Uncharted territory: Research on the socialization of special education teachers. *Teacher Education and Special Education*, *15*(2), 133–47. https://doi.org/http://dx.doi.org/10.1177/088840649201500208

Qu, S. Q., & Dumay, J. (2011) The qualitative research interview. *Qualitative Research in Accounting & Management*, *8*(3), 238-264. https://doi.org/10.1108/11766091111162070

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, *28*, 4–13. https://doi.org/http://dx.doi.org/10.1002/bs.3830280103

Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education*, *26*(7), 1372–1380. https://doi.org/10.1016/j.tate.2010.03.008

Rawlins, P., Brandon, J., Chapman, J., Leach, L., Neutze, G., Scott, A., & Zepke, N. (2005). Standards-based assessment in the senior secondary school: A review of the literature. *New Zealand Journal of Teachers' Work*, *2*(2), 107–115.

Remesal, A. (2011). Primary and secondary teachers' conceptions of assessment: A qualitative study. *Teaching and Teacher Education*, *27*(2), 472–482. https://doi.org/10.1016/j.tate.2010.09.017

Renwick, M., & Vise, J. (1993). *Windows on teacher education*. Wellington, New Zealand: New Zealand Council of Educational Research.

Renzaglia, A., Hutchins, M., & Lee, S. (1997). The impact of teacher education on the beliefs, attitudes and dispositions and pre-service special educators.

Teacher Education and Special Education, *4*, 360–377.
https://doi.org/http://dx.doi.org/10.1177/088840649702000406

Rice, J. K. (2003). *Teacher quality: Understanding the effectiveness of teacher attributes.* Washington, DC: Economic Policy Institute.

Richardson, V. (2003). Preservice teachers' beliefs. In J. Raths & A. McAnnich (Eds.), *Teacher beliefs and classroom performance: The impact of teacher education* (pp. 1–22). Greenwich, CT: Information Age Pub.

Rinchen, S., Ritchie, S. M., & Bellocchi, A. (2016). Emotional climate of a pre-service science teacher education class in Bhutan. *Cultural Studies of Science Education*, *11*(3), 603–628. https://doi.org/10.1007/s11422-014-9658-0

Ritchie, S. M., Tobin, K., Hudson, P., Roth, W.-M., & Mergard, V. (2011). Reproducing successful rituals in bad times: Exploring emotional interactions of a new science teacher. *Science Education*, *95*(4), 745–765. https://doi.org/10.1002/sce.20440

Robinson, V., & Timperley, H. (2000). The link between accountability and improvement: The case of reporting to parents. *Peabody Journal of Education*, *75*(4), 66–89.

Rogers, G. (2011). Learning-to-learn and learning-to-teach: The impact of disciplinary subject study on student-teachers' professional identity. *Journal of Curriculum Studies*, *43*(2), 249–268. https://doi.org/http://dx.doi.org/10.1080/00220272.2010.521262

Rogoff, B. (2008). Observing sociocultural activity on three planes: Participatory appropriation, guided participation and apprenticeship. In K. Hall, P. Murphy, & J. Soler (Eds.), *Pedagogy and practice: Culture and identities* (pp. 58–74). London, England: Sage Publications.

Rozenszajn, R., & Yarden, A. (2014). Expansion of biology teachers' pedagogical content knowledge (PCK) during a long-term professional development program. *Research in Science Education*, *44*(1), 189–213. https://doi.org/10.1007/s11165-013-9378-6

Rutherford, F. J. (2003). Reflections on assessment. In J. M. Atkin & J. E. Coffey (Eds.), *Everyday assessment in the science classroom* (pp. 147–158). Arlington, VA: National Science Teachers Association.

Sadler, D. R. (1987). Specifying and Promulgating Achievement Standards. *Oxford Review of Education*, *13*(2), 191–209.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144. https://doi.org/10.1007/BF00117714

Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education*, *5*(1), 77–84.

Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, *34*(2), 159–179. https://doi.org/10.1080/02602930801956059

Sarantakos, S. (1998). *Social research* (2nd ed.). South Yarra, Australia: Macmillan.

Scardamalia, M., Bransford, J. D., Kozma, R., & Quellmalz, E. (2012). New assessments and environments for knowledge building. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 231–300). Dordrecht, The Netherlands: Springer.

Schildkamp, K., & Ehren, M. (2013). From 'Intuition'- to 'Data'-based decision making in Dutch secondary schools? In K. Schildkamp, M. K. Lai, & L. Earl (Eds.), *Data-based decision making in education: Challenges and opportunities* (Vol. 17, pp. 49–68). Dordrecht, The Netherlands: Springer.

Schildkamp, K., & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education*, *26*, 482–496.

Schildkamp, K., Lai, M. K., & Earl, L. (Eds.). (2013). *Data-based decision making in education: Challenges and opportunities*. Dordrecht, The Netherlands: Springer.

Schneider, M. C., Egan, K. L., & Julian, M. W. (2013). Classroom assessment in the context of high-stakes testing. In J.H. McMillan (Ed.), *Sage Handbook of classroom assessment* (pp. 55–70). Los Angeles, CA: Sage Publications.

Schneider, R. M., & Plasman, K. (2011). Science teacher learning progressions: A review of science teachers' pedagogical content knowledge development. *Review of Educational Research*, *81*(4), 530–565. https://doi.org/10.3102/0034654311423382

Schwandt, T. A. (1994). Constructivist, interpretivist approaches to human inquiry. In N. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research*. Thousand Oaks, CA: Sage Publications.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation*. Chicago, IL: Rand McNally.

Seah, W. T. (2002). The perception of, and interaction with, value differences by immigrant teachers of mathematics in two Australian secondary classrooms. *Journal of Intercultural Studies*, *23*(2), 189–210. http://dx.doi.org/10.1080/07256860220151087

Semin, G. R., & Smith, E. R. (2013). Socially situated cognition in perspective. *Social Cognition*, *31*(2), 125–146. https://doi.org/http://dx.doi.org.ezproxy.waikato.ac.nz/101521soco201331 2125

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing one. *Educational Researcher*, *27*(2), 4–13.

Shapira-Lishchinsky, O. (2011). Teachers' critical incidents: Ethical dilemmas in teaching practice. *Teaching and Teacher Education*, *27*(3), 648–656. https://doi.org/10.1016/j.tate.2010.11.003

Shapiro, S. (2010). Revisiting the teachers' lounge: Reflections on emotional experience and teacher identity. *Teaching and Teacher Education*, *26*(3), 616–621. https://doi.org/10.1016/j.tate.2009.09.009

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, *29*(7), 4–14. https://doi.org/10.2307/1176145

Shepard, L. A. (2006). Classroom assessment. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 623-646). Westport, CT: Praeger Publishers.

Shepard, L. A., Hammerness, K., Darling-Hammond, L., Rust, F., Baratz Snowden, J., Gordon, E., … Pacheco, A. (2012). Assessment. In L. Darling-Hammond & J. D. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 275–326). San Francisco, CA: Jossey-Bass.

Shoffner, M. (2009). The place of the personal: Exploring the affective domain through reflection in teacher preparation. *Teaching and Teacher Education*, *25*(6), 783–789. https://doi.org/10.1016/j.tate.2008.11.012

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14. https://doi.org/10.2307/1175860

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reforms. *Harvard Educational Review*, *57*, 1–22. https://doi.org/http://dx.doi.org/10.17763/haer.57.1.j463w79r56455411

Shulman, L. S. (2015). PCK: Its genesis and exodus. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 3–13). New York, NY: Routledge.

Shulman, L. S., & Shulman, J. H. (2004). How and what teachers learn: A shifting perspective. *Journal of Curriculum Studies*, *36*(2), 257–271. https://doi.org/http://dx.doi.org/10.1080/0022027032000148298

Shulruf, B., Hattie, J., & Tumen, S. (2010). New Zealand's standard-based assessment for secondary schools (NCEA): Implications for policy makers. *Asia Pacific Journal of Education*, *30*(2), 141–165. https://doi.org/10.1080/02188791003721598

Shute, V. J., & Kim, Y. J. (2014). Formative and stealth assessment. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 311–321). New York, NY: Springer. https://doi.org/10.1007/978-1-4614-3185-5_25

Siegel, M. A. (2014). Developing preservice teachers' expertise in equitable assessment for English learners. *Journal of Science Teacher Education, 25,* 289-308. DOI: 10.1007/s10972-013-9365-9

Siegel, M. A., & Wissehr, C. (2011). Preparing for the plunge: Preservice teachers' assessment literacy. *Journal of Science Teacher Education*, *22*(4), 371–391. https://doi.org/10.1007/s10972-011-9231-6

Simon, S., & Campbell, S. (2012). Teacher learning and professional development in science education. In B. J. Fraser, K. G. Tobin, & C. J. McRobbie

(Eds.), *Second international handbook of science education Vol 1* (pp. 307–322). Dordrecht, The Netherlands: Springer.

Siry, C., & Brendel, M. (2016). The inseparable role of emotions in the teaching and learning of primary school science. *Cultural Studies of Science Education*, *11*(3), 803–815. https://doi.org/10.1007/s11422-016-9781-1

Sjolie, E. (2014). The role of theory in teacher education: Reconsidered from a student teacher perspective. *Journal of Curriculum Studies*, *46*(6), 729–750. https://doi.org/http://dx.doi.org/10.1080/00220272.2013.871754

Sluijsmans, D. M. A., & Struyven, K. (2014). Quality assurance in assessment: An introduction to this special issue. *Studies in Educational Evaluation*, *43*, 1–4. https://doi.org/10.1016/j.stueduc.2014.08.003

Smaill, E. (2013). Moderating New Zealand's national standards: Teacher learning and assessment outcomes. *Assessment in Education: Principles, Policy & Practice*, *20*(3), 250–265. https://doi.org/10.1080/0969594X.2012.696241

Smith, D. C., & Jang, S. (2011). Pathways in learning to teach elementary science: Navigating contexts, roles, affordances and constraints. *Journal of Science Teacher Education*, *22*(8), 745–768. https://doi.org/10.1007/s10972-011-9257-9

Smith, D. L. (1996). Teacher decision making in the adult ESL classroom. In D. Freeman & J. Richards (Eds.), *Teacher learning in language teaching* (pp. 197–216). Cambridge, NY: Cambridge University Press.

Smith, J. K. (2003). Reconsidering reliability in classroom assessment and grading. *Educational Measurement: Issues and Practice*, *22*(4), 26–33. https://doi.org/10.1111/j.1745-3992.2003.tb00141.x

Smith, K. (2010). Assessment: Complex concept and complex practice. *Assessment Matters*, *2*, 28–42.

Smith, K., & Lev-Ari, L. (2005). The place of the practicum in pre-service teacher education: The voice of the students. *Asia-Pacific Journal of Teacher Education*, *33*(3), 289–302. https://doi.org/10.1080/13598660500286333

Smith, L. F., Corkery, G., Buckley, J., & Calvert, A. (2013). Changes in secondary school preservice teachers' concerns about teaching in New Zealand. *Journal of Teacher Education*, *64*(1), 60–74. https://doi.org/10.1177/0022487112449019

Smith, L. F., Hill, M. F., Cowie, B., & Gilmore, A. (2014). Preparing teachers to use the enabling power of assessment. In Claire Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning* (pp. 303–323). Dordrecht, The Netherlands: Springer.

Soini, T., Pyhältö, K., & Pietarinen, J. (2010). Pedagogical well-being: Reflecting learning and well-being in teachers' work. *Teachers and Teaching*, *16*(6), 735–751. https://doi.org/10.1080/13540602.2010.517690

Spangler, G., Pekrun, R., Kramer, K., & Hofmann, H. (2002). Students' emotions, physiological reactions, and coping in academic exams. *Anxiety, Stress, & Coping*, *15*(4), 413–432. https://doi.org/10.1080/1061580021000056555

Starkey, L., Yates, A., Meyer, L. H., Hall, C., Taylor, M., Stevens, S., & Toia, R. (2009). Professional development design: Embedding educational reform in New Zealand. *Teaching and Teacher Education*, *25*(1), 181–189. https://doi.org/10.1016/j.tate.2008.08.007

Steiner Engelsen, K., & Smith, K. (2014). Assessment literacy. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning* (Vol. 1, pp. 91–108). Dordrecht, The Netherlands: Springer.

Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, *72*(7), 534–539.

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, *77*(3), 238–246.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment FOR learning. *Phi Delta Kappan*, *83*(10), 758–766. https://doi.org/http://dx.doi.org/10.1177/003172170208301010

Stiggins, R. J. (2007). Conquering the formative assessment frontier. In J.H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 8–28). New York, NY: Teachers College Press.

Stiggins, R. J. (2014). Improve assessment literacy outside of schools too: teaching and assessment have become separated, which has kept teachers from developing the assessment skills they need to truly enhance learning. *Phi Delta Kappan*, *96*(2), 67–72. https://doi.org/10.1177/0031721714553413

Stiggins, R. J., Arter, J. A., Chappuis, J., & Chappuis, S. (2004). *Classroom assessment for student learning: Doing it right, using it well*. Portland, OR: Assessment Training Institute.

Stoate, G. A. (2013). *The dialogic aspects of Mantle of the Expert pedagogy used to teach devising at NCEA Level 2 in a Year 12 classroom 'I don't think it's about credits- definitely not about credits'* (Master of Education Thesis, University of Waikato, Hamilton, New Zealand). Retrieved from http://researchcommons.waikato.ac.nz/handle/10289/7963

Stobart, G. (2001). The validity of national curriculum assessment. *British Journal of Educational Studies*, *49*(1), 26–39.

Stobart, G. (2005). Fairness in multicultural assessment systems. *Assessment in Education: Principles, Policy & Practice*, *12*(3), 275–287. https://doi.org/10.1080/09695940500337249

Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. New York, NY: Routledge.

Strike, K. A. (1990). Teaching ethics to teachers: What the curriculum should be about. *Teaching and Teacher Education*, *6*(1), 47–53. https://doi.org/10.1016/0742-051X(90)90006-Q

Struyven, K., Dochy, F., & Janssens, S. (2008). The effects of hands-on experience on students' preferences for assessment methods. *Journal of Teacher Education, 59*(1), 69-88. DOI: 10.1177/0022487107311335

Stuart, C., & Thurlow, D. (2000). Making it their own: Pre-service teachers' experiences, beliefs and classroom practices. *Journal of Teacher Education*, *51*(1), 113–121. https://doi.org/http://dx.doi.org/10.1177/0022487100051000205

Sutton, R. (1992). *Assessment: A framework for teachers*. London, England: Routledge.

Sutton, R. E., & Wheatley, K. F. (2003). Teachers' emotions and teaching: A review of the literature and directions for future research. *Educational Psychology Review*, *15*(4), 327–358. https://doi.org/10.1023/A:1026131715856

Suurtamm, C., & Koch, M. J. (2014). Navigating dilemmas in transforming assessment practices: experiences of mathematics teachers in Ontario, Canada. *Educational Assessment, Evaluation and Accountability*, *26*(3), 263–287. https://doi.org/10.1007/s11092-014-9195-0

Taber, K. S., Riga, F., Brindley, S., Winterbottom, M., Finney, J., & Fisher, L. G. (2011). Formative conceptions of assessment: Trainee teachers' thinking about assessment issues in English secondary schools. *Teacher Development*, *15*(2), 171–186. https://doi.org/10.1080/13664530.2011.571500

Tabor, P. D. (2008). *An investigation of instruction in two-digit addition and subtraction using a classroom teaching experiment methodology, design research, and multilevel modelling* (Doctoral thesis, Southern Cross University, Lismore, Australia). Retrieved from http://epubs.scu.edu.au/theses/68/

Tamir, P. (1988). Subject matter and related pedagogical knowledge in teacher education. *Teaching and Teacher Education*, *4*(2), 99–110. https://doi.org/10.1016/0742-051X(88)90011-X

Taras, M. (2005). Assessment – summative and formative – some theoretical reflections. *British Journal of Educational Studies*, *53*(4), 466–478. https://doi.org/10.1111/j.1467-8527.2005.00307.x

Taras, M. (2009). Summative assessment: The missing link for formative assessment. *Journal of Further and Higher Education*, *33*(1), 57–69. https://doi.org/10.1080/03098770802638671

Te Kete Ipurangi. (n.d.). Assessment literacy. Retrieved from http://assessment.tki.org.nz/Assessment-in-the-classroom/Assessment-for-learning-in-practice/Assessment-literacy

TeachNZ. (2006). Advanced classroom expertise teacher recognition. Retrieved from https://www. teachnz.govt.nz/home/advanced-classroom-expertise-teacher-recognition/

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127–146). New York, NY: Macmillan.

Thompson, G. (2010). Acting, accidents and performativity: Challenging the hegemonic good student in secondary schools. *British Journal of Sociology of Education*, *31*(4), 413–430. https://doi.org/10.1080/01425692.2010.484919

Thrupp, M., & Easter, A. (2012). *Research, analysis and insight into national standards (RAINS) project* (No. 1). Wellington, New Zealand: The New Zealand Educational Institute Te Riu Roa.

Thrupp, M. (2013). National standards for student achievement: Is New Zealand's idiosyncratic approach any better? *Australian Journal of Language and Literacy*, *36*(2), 99.

Tierney, R. D. (2013). Fairness in classroom assessment. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 125–144). Thousand Oaks, CA: Sage Publications.

Tierney, R. D. (2014). Fairness as a multifaceted quality in classroom assessment. *Studies in Educational Evaluation*, *43*, 55–69. https://doi.org/10.1016/j.stueduc.2013.12.003

Tierney, R. D., Simon, M., & Charland, J. (2011). Being fair: Teachers' interpretations of principles for standards-based grading. *The Educational Forum*, *75*, 210–227. http://dx.doi.org/10.1080/00131725.2011.577669

Timperley, H. (2009). Using assessment data for improving teaching practice. *Australian College of Educators*, *8*(3), 21–27.

Timperley, H., & Parr, J. (2004). *Using evidence in teaching practice: Implications for professional learning*. Auckland, New Zealand: Hodder Moa Beckett.

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education: Principles, Policy & Practice*, *14*(3), 281–294. https://doi.org/10.1080/09695940701591867

Torrance, H. (2015). Blaming the victim: Assessment, examinations, and the responsibilisation of students and teachers in neo-liberal governance. *Discourse: Studies in the Cultural Politics of Education*. 38(1), 83-96. https://doi.org/10.1080/01596306.2015.1104854

Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom.* Florence, KY: Taylor & Francis.

Trumbell, D. J. (1999). *The new science teacher*. New York, NY: Teachers College Press.

Tsui, A. B. M. (2009). Distinctive qualities of expert teachers. *Teachers and Teaching*, *15*(4), 421–439. https://doi.org/10.1080/13540600903057179

Unger, J. (2013). Flex your school's data muscles: Leadership strategies strengthen data's impact. *Learning Forward*, *34*(4), 50–54,

Ussher, B., & Earl, K. (2010). 'Summative' and 'Formative': Confused by the assessment terms? *New Zealand Journal of Teachers' Work*, *7*(1), 53–63.

Van Driel, J. H., & Berry, A. (2012). Teacher professional development focusing on pedagogical content knowledge. *Educational Researcher*, *41*(1), 26–28. https://doi.org/10.3102/0013189X11431010

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT Press.

Veenman, S. (1984). Perceived problems of beginning teachers. *Review of Educational Research*, *54*(2), 143–178. https://doi.org/10.2307/1170301

Venville, G., Sheffield, R., Rennie, L. J., & Wallace, J. (2008). The writing on the wall: Classroom context, curriculum implementation, and student learning in integrated, community-based science projects. *Journal of Research in Science Teaching*, *45*(8), 857–880. https://doi.org/10.1002/tea.20245

Vogl, E., & Pekrun, R. (2016). Emotions that matter to achievement. In G.T.L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 111–128). New York, NY: Routledge.

Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, *30*(3), 749–770. https://doi.org/10.2307/20466661

Vonk, J. H. C. (1993). Mentoring beginning teachers: Development of a knowledge base for mentors. Paper presented at the Annual meeting of the American Educational Research Association, Atlanta, Georgia, USA.

Vygotsky, L. (1978). *Mind and society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wallace, J., & Loughran, J. (2012). Science teacher learning. In B. J. Fraser, K. G. Tobin, & C. J. McRobbie (Eds.), *Second international handbook of science education* (pp. 295–306). Dordrecht, The Netherlands: Springer.

Wang, J., Kao, H., & Lin, S. (2010). Preservice teachers' initial conceptions about assessment of science learning: The coherence with their views of learning science. *Teaching and Teacher Education*, *26*(3), 522–529. https://doi.org/http://dx.doi.org/10.1016/j.tate.2009.06.014

Wang, T.-H., Wang, K.-H., & Huang, S.-C. (2008). Designing a web-based assessment environment for improving pre-service teacher assessment literacy. *Computers & Education*, *51*(1), 448–462. https://doi.org/10.1016/j.compedu.2007.06.010

Watkins, M. (2007). Thwarting desire: discursive constraint and pedagogic practice. *International Journal of Qualitative Studies in Education*, *20*(3), 301–318. https://doi.org/10.1080/09518390701281900

Weeden, P., Winter, J., & Broadfoot, P. (2002). *Assessment: What's in it for schools?* London, England: RoutledgeFalmer.

Weinstein, C. S. (1998). 'I want to be nice, but I have to be mean': Exploring prospective teachers' conceptions of caring and order. *Teaching and Teacher Education*, *14*(2), 153–163. https://doi.org/10.1016/S0742-051X(97)00034-6

Whitehead, D. (2007). Literacy assessment practices: Moving from standardised to ecologically valid assessments in secondary schools. *Language and Education*, *21*(5), 434–452. https://doi.org/10.2167/le801.0

Wiliam, D. (2006). Assessment: Learning communities can use IT to engineer a bridge connecting teaching and learning. *Journal of Staff Development*, *27*(1), 16–20.

Wiliam, D., & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal*, *22*(5), 537–548.

Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–87). Mahwah, NJ: Erlbaum.

Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualising teachers' assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, *40*(2), 241–256. https://doi.org/10.1007/s13384-013-0089-9

Wilson, E., & Demetriou, H. (2007). New teacher learning: Substantive knowledge and contextual factors. *The Curriculum Journal*, *18*(3), 213–229. https://doi.org/10.1080/09585170701589710

Winograd, K. (2003). The functions of teacher emotions: The good, the bad, and the ugly. *Teachers College Record*, *105*, 1641–1673. http://dx.doi.org/10.1046/j.1467-9620.2003.00304.x

Winterbottom, M., Brindley, S., Taber, K. S., Fisher, L. G., Finney, J., & Riga, F. (2008). Conceptions of assessment: Trainee teachers' practice and values. *Curriculum Journal*, *19*(3), 193–213. https://doi.org/10.1080/09585170802357504

Winterbottom, M., Taber, K. S., Brindley, S., Fisher, L. G., Finney, J., & Riga, F. (2008). Understanding differences in trainee teachers' values and practice in relation to assessment. *Teacher Development*, *12*(1), 15–35. https://doi.org/10.1080/13664530701827723

Woods, P., & Jeffrey, B. (2002). The reconstruction of primary teachers' identities. *British Journal of Sociology of Education*, *23*(1), 89–106. http://dx.doi.org/10.1080/01425690120102872

Wyatt-Smith, C., & Gunn, S. (2009). Towards theorizing assessment as critical inquiry. In C. Wyatt-Smith & J. J. Cumming (Eds.), *Educational*

assessment in the 21st Century: Connecting theory and practice (pp. 83–102). Dordrecht, The Netherlands: Springer.

Wyatt-Smith, C., Klenowski, V., & Colbert, P. (2014). Assessment understood as enabling. In Claire Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning* (Vol. 1, pp. 1–22). Heidelberg, Germany: Springer.

Wyatt-Smith, C., Klenowski, V., & Gunn, S. (2010). The centrality of teachers' judgements practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, Policy & Practice*, *17*(1), 59–75. https://doi.org/http://dx.doi.org/10.1080/09695940903565610

Wylie, C. (2013). *Secondary schools in 2012: Main findings from the NZCER national survey*. Wellington, New Zealand: New Zealand Council for Educational Research.

Wylie, C., & Bonne, L. (2016). *Secondary schools in 2015: Findings from the NZCER national survey*. Wellington, New Zealand: New Zealand Council for Educational Research.

Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly*, *43*(3), 493–513.

Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149–162. https://doi.org/10.1016/j.tate.2016.05.010

Yates, A., & Johnston, M. (2017). The impact of school-based assessment for qualifications on teachers' conceptions of assessment. *Assessment in Education: Principles, Policy & Practice*, 1–17. https://doi.org/10.1080/0969594X.2017.1295020

Yilmaz-Tuzun, O. (2007). Preservice elementary teachers' beliefs about science teaching. *Journal of Science Teacher Education*, *19*(2), 183–204. https://doi.org/10.1007/s10972-007-9084-1

Yin, R. (1988). *Case study research: design and methods*. Thousand Oaks, CA: Sage Publications.

Youngs, P., & Bird, T. (2010). Using embedded assessments to promote pedagogical reasoning among secondary teaching candidates. *Teaching and Teacher Education*, *26*(2), 185–198. https://doi.org/http://dx.doi.org/10.1016/j.tate.2009.03.011

Yung, B. H.-W. (2002). Same assessment, different practice: Professional consciousness as a determinant of teachers' practice in a school-based assessment scheme. *Assessment in Education: Principles, Policy & Practice*, *9*(1), 97–117. https://doi.org/10.1080/09695940220119210

Zeidner, M. (1995). Adaptive coping with test situations: A review of the literature. *Educational Psychologist*, *30*(3), 123–133. https://doi.org/10.1207/s15326985ep3003_3

Zembylas, M. (2002). Constructing genealogies of teachers' emotions in science teaching. *Journal of Research in Science Teaching*, *39*, 79–103. https://dx.doi.org/10.1002/tea.10010

Zembylas, M. (2004). The emotional characteristics of teaching: An ethnographic study of one teacher. *Teaching and Teacher Education*, *20*, 185–201. https://doi.org/10.1016/j.tate.2003.09.008

Zembylas, M. (2007). Emotional ecology: The intersection of emotional knowledge and pedagogical content knowledge in teaching. *Teaching and Teacher Education*, *23*(4), 355–367. https://doi.org/10.1016/j.tate.2006.12.002

Zembylas, M. (2014). Making sense of the complex entanglement between emotion and pedagogy: Contributions of the affective turn. *Cultural Studies of Science Education*, *11*(3), 539–550. https://doi.org/10.1007/s11422-014-9623-y

Zembylas, M., & Barker, H. (2002). Preservice teacher attitudes and emotions: Individual spaces, community conversations and transformations *Research in Science Education*, *32*, 329–351. doi:10.1023/A:1020862000107

Appendices

Appendix A: Participant letter of invitation

26 February 2013

Dear members of TEMS726

Invitation to participate in research

Assessment process and practice is central to good teaching in New Zealand secondary schools, and has been highlighted with the move to standards-based assessment. My experience as a teacher, a school leader, and NZQA moderator shows me that teachers need to have capability in assessment in order to be able to operate effectively in the classroom. This capability can be described as assessment know-how and motivation. I am particularly interested in exploring how beginning teachers develop assessment capability through their initial teacher education and their first year of teaching. I would like to work with you during this section of your career so that I can better understand the influences and factors at play as you develop your assessment capability.

This research will contribute to my doctoral thesis and will inform our work at the University of Waikato, as we continually strive to provide teacher education programmes that best prepare teachers for their work in secondary schools. This research is being supervised by Associate Professor Beverley Bell and Associate Professor Bronwen Cowie, both based in the Faculty of Education at the University of Waikato.

I am a science teacher myself and have been involved in assessment both as a teacher and in various roles I have held with NZQA. I moved to work at the University just 3 years ago, so the challenge of using meaningful assessment for secondary students is still very fresh in my mind.

I appreciate that the decision to participate in this research must be a considered one. Please find attached an information sheet that outlines the research

questions, your involvement, ethical considerations and my personal commitment to you.

If you are interested in participating in this research, and I sincerely hope you are, I would appreciate it if you could read the attached information sheet and consent form. I will meet you in class in the near future and at that point we can discuss the research further and make arrangements for first meeting if you are willing to be involved in this project. As part of this research I will be observing some university classes you are involved in, only after receiving permission to do so from the lecturer and the class involved.

Thank you for considering this invitation. I look forward to working with you.

Kind regards

Frances Edwards

francese@waikato.ac.nz

Ph: 8384455 ext.6170

Appendix B: Participant consent form

Participant consent form

Research topic: The development of summative assessment capability: an exploration of the experiences of beginning secondary science teachers in New Zealand

Researcher: Frances Edwards

I have been fully informed as to the purpose, potential ethical concerns and processes involved in this research study and freely consent to participate. I understand that if I have any concerns regarding my involvement in this study, these can be addressed at any stage of the research process. I have read the information sheet, and understand the nature of the study. I have had the opportunity to ask questions and have them answered.

I understand that I am free to withdraw from this study at any time without given a reason, until 1 December 2013 for Phase 1 participants, and 1 July 2014 for Phase 2 participants. Should I choose to withdraw I will do this in writing and the researcher will destroy all records pertaining to me.

I understand that this study is approximately 18 months long, and I agree to participate in the completion of 2 questionnaires and 4 confidential semi-structured interviews as a Phase 1 participant. Should I be a participant in Phase 2 of the study ie I secure a teaching position for 2014, I agree to be involved in one further interview. I am aware that the interviews will be recorded and transcribed. I will receive copies of the transcripts of interviews for verification. I will have the opportunity to amend and make comment on this data before the researcher uses it in writing her doctoral thesis or in any publications. I understand that all non-identifying research data will be kept indefinitely and that records containing my personal information will be destroyed 5 years after the completion of this project. I agree to provide copies of artefacts which display information about my developing understanding of assessment, and that these may include assessments from courses and practicum materials, and my diary.

I am aware that from time to time the researcher may be a non-participant observer in teacher education classes I attend in 2013. Before any observation takes place I

understand that consent will be sought for and gained from the lecturer and the class involved.

I understand that I retain ownership of the raw data recorded and transcribed, and that at the conclusion of this research project I will receive an abstract and electronic link to the completed thesis.

I understand that my lecturers and programme coordinator will not be informed about my participation or non-participation in this research project.

I am aware that at any time during this research project I can express concerns to the research and/or her supervisors, and if I do not receive a satisfactory outcome I can make a formal complaint to the Faculty of Education Ethics Committee as described in Section 24 of the *Ethical Conduct in Human Research and Related Activities Regulations*(2008).

Signed...

Name:...

Address:..

Phone number:....................................... Email address:..

Date:......................................

Appendix C: Information sheet for participants

Information Sheet

This sheet outlines the proposed research and contains information that will help you make an informed decision about your participation.

Research topic:

The development of summative assessment capability: an exploration of the experiences of beginning secondary science teachers in New Zealand

In seeking to understand the development of summative assessment capability in beginning science teachers this study poses three broad questions:

   a)   What are the characteristics of summative assessment capability of beginning secondary science teachers over their ITE year and as they start teaching?
   b)   In what ways do beginning secondary science teachers' summative assessment thinking and capabilities develop/change over their ITE year and as they start teaching?
   c)   What contributes to change in the assessment capability of beginning secondary science teachers over their ITE year and as they start teaching?

As you can see, I would like to find out how your thinking and practice develops over time, and what assessment capability looks like for beginner teachers. This research is descriptive and will involve conversations around your development as a teacher. You may well have questions about this subject that are important to you and these can be included in our conversations.

Research activities:

Phase 1

Over a period of 10 months I will invite you to be involved in a number of conversations where we can talk about your experiences and development of assessment capability. I anticipate we will meet for 4 such conversations (before and after each practicum), unless you wish to meet more frequently. I would also value you completing a brief questionnaire at the beginning and the end of your year of study (February and November), and keeping a diary in which ideas and reflections about assessment are jotted down at your discretion. I will ask for copies of materials you produce (eg assignments, lesson plans and practicum folders), which demonstrate your developing

assessment capability. These will be photocopied and returned. I will also be present from time to time as a non-participant observer in teacher education classes you attend.

Phase 2

If you have secured a teaching position for 2014 I would like to invite you to be involved in one more interview with me (June 2014).

In both phases all of our conversations will be in the form of semi-structured interviews of approximately 50 minutes duration. I would like to audiotape these interviews with a digital recorder, with your permission, and make field notes as we talk.

Your rights as a participant

The decision to participate in this research is yours alone. If you decide that you do want to be involved, you may, without explanation:

1. decline to answer any questions at any time, and/or

2. withdraw any information provided at any time prior to 1 December 2013 for Phase 1 participants, and/or

3. withdraw any information provided at any time prior to 1 July 2014 for Phase 2 participants

4. withdraw totally from the research study at any time prior to 1 December 2013 for Phase 1 participants, or prior to 1 July 2014 for Phase 2 participants.

You will have open access to all your interview transcripts, copies of which will be provided to you. There may be times when you may wish to censor some of your responses. I will not make any changes to transcripts/conversation field notes without your express knowledge and consent.

Ethical considerations:

1. Confidentiality

In order to maintain confidentiality your involvement in this project will not be disclosed to any person. Your identity will be kept confidential and a pseudonym or your middle name will be used. I will avoid naming the University of Waikato or the school in which you are employed in 2014.

2. Potential for harm

If any information is disclosed by you that could potentially harm yourself or others I have a responsibility to draw your attention to this and not use or edit material accordingly. When references are inadvertently made to others, their identity will remain confidential. Nothing you reveal to me will be made available to staff involved in the secondary graduate programme. The only exception is that a brief collective summary of findings will be provided to staff after the 2013 programme is complete. This will be summary in nature, so will not involve the disclosure of any individual names and data.

3. Conflicts of interest

I teach at the University of Waikato, but I am not involved in any way in the teaching or assessment of courses you are involved in, nor will I be involved in visiting you on practicum, or involved in any meetings where there may be discussion regarding your progress within your enrolled programme. I have not identified any conflict of interest but if either of us become aware of such conflict it is important that this is declared and that my supervisors are informed. If a simple solution cannot be found the Faculty of Education Ethics Committee may be asked for advice. The well-being of participants takes precedence over the research.

4. Procedures for handling information and materials produced in the course of this research

Security

All electronic data (digital interview recordings, interview transcripts and pdf scans of field notes and diary entries, questionnaire analysis) will be stored on my work computer with back-up copies kept on a central university server and on my portable hard drive. Access to all of these is password protected. All hard copies of field notes, questionnaires, and artefacts provided by you will be kept in a locked filing cabinet in my university office. In accordance with the *Ethical Conduct in Human Research and Related Activities Regulations* (2008) all non-identifying data (transcripts, field notes, questionnaires and photocopied artefacts) used for my doctoral thesis will be archived indefinitely. Consent forms containing your personal information will be destroyed five years after the completion of the study.

Access to information

As the researcher I will have sole access to personal information on your consent form. Access to the raw data will be limited to you and me, although either of my supervisors may wish to audit aspects of the data analysis process. You will receive copies of the interview transcripts and a thesis abstract which includes a link to the Australasian Digital Thesis database that hosts all completed theses. An electronic copy of the thesis will become widely available, as the University of Waikato requires that a digital copy of Doctoral theses will be lodged permanently in the University's digital repository: Research Commons.

Ownership of data or materials produced

You own all of the responses you have made to questions in interviews and questionnaires, and artefacts from your course work and diary entries. The research findings, resultant thesis and any subsequent publications and presentations belong to me.

Copyright

This research will conform to the University of Waikato Copyright Guidelines 2nd ed (2009). These are available at
http://www.waikato.ac.nz/copyright/uow_copyright_guidelines_April2009.pdf The copyright of the doctoral thesis and sole-authored scholarly publications and presentations belongs to the researcher.

5. Procedure for resolution of disputes

This research is bound by the *Conduct in Human Research and Related Activities Regulations*(2008) which can be found at this link
http://calendar.waikato.ac.nz/assessment/ethicalConduct.html and approved by the Faculty of Education Research Ethics Committee. If you are concerned over any process or procedure involved in this research you may wish to contact my supervisors directly and inform them of the concern. They will negotiate a process for mediation if this is required, and if this is not completed to your satisfaction, you are able to submit a formal complaint (see Section 24 of the regulations mentioned above).

Supervisors contact details

Associate Professor Beverley Bell: bbell@waikato.ac.nz Ph 8384455 ext 4101

Associate Professor Bronwen Cowie: bcowie@waikato.ac.nz Ph 8384455 ext 4987

If you have any further questions about this research, please ask.

Kind regards

Frances Edwards

francese@waikato.ac.nz

Appendix D: Interview schedule

The development of summative assessment capability: an exploration of the experiences of beginning secondary science teachers in New Zealand

Interview 1 (Pre-practicum 1)

1. Tell me a bit about your past experiences – what drew you into teaching as a career?
2. What is your understanding of assessment? Of summative assessment?
3. What purpose/s do you think assessment serves?
4. How is assessment used in secondary classrooms?
5. How, if at all, have you used assessment as a teacher? Examples?
6. What have you learnt about assessment that has surprised you so far?
7. How "assessment capable" do you feel at the moment? Explain.
8. What do you need to learn about assessment? What skills do you bring to assessment?

Interview 2 (Post-practicum 1)

1. What is your understanding of assessment? Of summative assessment?
2. What purpose/s do you think assessment serves?
3. How is assessment used in secondary classrooms?
4. How, if at all, have you used assessment as a teacher? Examples?
5. What have you learnt about assessment that has surprised you so far?
6. What have you learnt about assessment from your practicum experience?
7. How has your practicum experience helped in your development as a teacher with respect to summative assessment?
8. What have you learnt about assessment from your university course experience?
9. How has your university course experience helped in your development as a teacher with respect to summative assessment?
10. How "assessment capable" do you feel at the moment? Explain.
11. What do you need to learn about assessment?

The pre and post-practicum 2 interviews were based on the above schedule with supplementary questions based on the responses of individual participants in earlier interviews. Teachers were also asked to talk about the artefacts they provided where applicable, and how these demonstrated their assessment literacy.

Appendix E: Summary of categories that resulted from data analysis.

Categories	Condensed Code	Contributing codes
Knowledge of assessment	Ability to define assessment	Definition of assessment
	Knowledge of purposes of summative assessment	Knowledge of assessment purposes
	Knowledge of what to assess	Knowledge of subject content and assessment
		Knowledge of what to assess
	Knowledge of strategies and design	Strategies for assessment
		Designing assessment tasks
	Knowledge of assessment interpretation	Data literacy
		Assessment leading to action
		Making judgements – marking
Understanding the context for assessment	NCEA assessment	NCEA assessment
	Preparing students for standards-based assessment	'Teaching to the test'
		Preparing students for summative assessment
	Using summative assessment formatively	Learning about assessment at university
		Ideas about learning
		Learning from practicum
		Assessment leads to action
Recognising the impact of assessment	Understanding assessment consequences	Benefits or effects of assessment for student
		Benefits or effects of assessment for teachers
		Students can monitor their progress
	Fairness	Definition of assessment including fairness
		Fairness
Self-analysis of personal development	Enablers and barriers	Assessment capability, teachers' self-assessment
		Motivation to teach
		Where ideas come from
	Personal responses as a teacher	Assessment capability, teachers' self-assessment
		Assessment is a reflection on and for the teacher

Appendix F: Sample quotes from categories that resulted from data analysis

Category	Sample quotes
Knowledge Of Assessment	
Ability to define assessment	Assessment is kind of looking at what you could do in regards like what you're expected to be able to do, like what you've been taught and what you can do from that. [Elisa, I1]
	I'd say assessment is a way to test the students' knowledge at a certain point. [Mary, I1]
Knowledge of purposes of summative assessment	Three types, the prior, and the formative, and the summative. Formative, I find I do that every lesson without even really realising, but I'm quite big on the formative. I guess like I was saying before, lots of quizzes and always asking questions. I guess that's my way of assessing to see if they are understanding the current lesson. [Bella, I2]
	…to gain an understanding of where they're at and how well the students are doing and the teachers are doing, and I guess from there you can make improvements. I guess that would be the overarching thing, so where to make changes to improve for the teacher, and how to make things improve for the students.[Benjamin, I1]
Knowledge of what to assess	I'd be able to assess their knowledge I guess but I wouldn't be able to assess so much their understanding or their thought processes [Benjamin, I2]
	When they had to do the structure of a flower in the test - the flower they were using was completely different. [Kate, I4]
Knowledge of strategies and design	So I had to make up my own template for Salmonella, and I broke it down into categories so broke it down to each aspect of MRS GREN. And then I broke down the second question into minor questions. [Bella, I4]
	For example in the senior biotech one, basically there are these two options, which one do you want to do? And we came up with a report supported with a poster, well it ended up being just supported with diagrams because it ended up being easier that way. [Susie, I4]
UNDERSTANDING THE CONTEXT FOR ASSESSMENT	
NCEA assessment	… and it was all to do with reading the exemplars and whatever they do for the exams. Because it was just practice exams, the NZQA thing [Elisa, I2]
	…especially when you've got that merit /excellence decision to make you can't just tick or cross, you have to do follow their thought processes through. [Benjamin, I2]
Preparing students for standards-based assessment	…got them to highlight the main points of the questions, the main words. And then for the words they highlight, they have to write any ideas that come to mind, just write the word next

	to it just all the different words that come to mind. Then they can incorporate that into their answer. [Bella, I4]
	I make sure they know the content, then I make sure they know how to go about answering the questions by breaking down the science question so they know how to answer it. [Wiremu, I4]
Using summative assessment formatively	well they were more of a evaluating the learning they got from the last lesson, to see if there was anything that they hadn't got, but I found they were pretty effective for that. [Ryan, I4]
	Actually doing some marking, and also getting some summative feedback from students… and then been able to pull back and draw on that for the next lesson, rather than just going on. [Mary, I5]

RECOGNISING THE IMPACT OF ASSESSMENT

Understanding assessment consequences	And then that almost became the panic mode like … we have to get credits, rather than like, you guys are all Year 11s and if you do your internals you're going to get heaps of credits. But they were just panicked that they weren't going to get them. [Mary, I2]
	…and the best anyone had done was 5/20. And then they'd done one at the end of the unit, and everyone got between 15 and 20 out of 20, and they were so proud of themselves. They went 'we actually learnt something'… that was cool. [Mary, I1]
Fairness	And I've always hated hypocrisy… so I think that will be good in making me fair with when to be flexible with assessment, and knowing when not to be too strict, because I think I will be able to do that, have that sort of flexibility, because I don't think I'd make compromises on fairness…[Wiremu, I1]
	Fairness, I guess, it's hard to define fairness because it's the same for everyone, but whether that is fair or not, I don't know….but maybe a better assessment might be to let students choose whether they want to give a written essay or give an oral presentation. I think that would be more fair. [Benjamin, I2]

Appendix G: Letter of ethical approval from the Faculty of Education Research Ethics Committee, University of Waikato

Dean's Office
Faculty of Education
Te Kura Toi Tangata
The University of Waikato
Private Bag 3105
Hamilton, New Zealand

Phone +64 7 838 4500
www.waikato.ac.nz

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

MEMORANDUM

To:	**Frances Edwards**
cc:	Associate Professor Beverley Bell
	Dr Elmarie Kotzé
From:	**Associate Professor Linda Mitchell**
	Chairperson, Research Ethics Committee
Date;	24 July 2012
Subject:	Supervised Postgraduate Research – Application for Ethical Approval (EDU068/12)

Thank you for submitting the amendments to your application for ethical approval for the research project:

**The development of summative assessment capabilty: an exploration
of the experiences of beginning secondary science teachers in New Zealand**

I am pleased to advise that your application has received ethical approval.

Please note that researchers are asked to consult with the Faculty's Research Ethics Committee in the first instance if any changes to the approved research design are proposed.

The Committee wishes you all the best with your research.

Associate Professor Linda Mitchell
Chairperson
Faculty of Education Research Ethics Committee

Appendix H: Individual Teachers' Assessment Literacy Development Across Ten Dimensions

Note: "Int" denotes Interview. Hence "Int 1" denotes Interview 1 and so forth. Levels of competence are 1-5, to denote Novice, Advanced Beginner, Competent, Proficient and Expert as explained in Table 5 in Chapter 4.

Teacher	Interview series				
Bella	**Int 1**	**Int 2**	**Int 3**	**Int 4**	**Int 5**
1. Ability to describe assessment	1	2	2	2	3
2. Knowledge of purposes of summative assessment	2	3	3	3	3
3. Knowledge of what to assess	1	3	3	3	3
4. Knowledge of assessment strategies and task design	1	2	2	2	2
5. Knowledge of assessment interpretation	1	3	3	3	3
6. NCEA assessment	1	2	2	2	3
7. Preparing students for standards-based assessment	0	0	0	2	3
8. Using summative assessment formatively	1	3	3	3	3
9. Understanding assessment consequences	2	2	2	3	3
10. Fairness	1	2	2	2	2
Benjamin	**Int 1**	**Int 2**	**Int 3**	**Int 4**	**Int 5**
1. Ability to describe assessment	1	2	2	2	2
2. Knowledge of purposes of summative assessment	1	2	2	3	3
3. Knowledge of what to assess	1	2	2	2	3
4. Knowledge of assessment strategies and task design	1	2	2	2	3
5. Knowledge of assessment interpretation	2	2	2	3	3
6. NCEA assessment	1	1	2	2	2
7. Preparing students for standards-based assessment	1	1	1	1	2
8. Using summative assessment formatively	2	3	3	3	3
9. Understanding assessment consequences	2	2	2	2	2
10. Fairness	1	2	2	2	2
Elisa	**Int 1**	**Int 2**	**Int 3**	**Int 4**	**Int 5**
1. Ability to describe assessment	1	2	2	2	3
2. Knowledge of purposes of summative assessment	1	2	3	3	3
3. Knowledge of what to assess	1	2	2	3	3
4. Knowledge of assessment strategies and design	1	2	2	2	3
5. Knowledge of assessment interpretation	1	2	2	2	3
6. NCEA assessment	1	2	2	3	3
7. Preparing students for standards-based assessment	0	2	2	3	3
8. Using summative assessment formatively	2	3	3	3	3
9. Understanding assessment consequences	1	2	2	3	3
10. Fairness	1	2	2	3	3
Kate	**Int 1**	**Int 2**	**Int 3**	**Int 4**	**Int 5**
1. Ability to describe assessment	1	2	2	3	3
2. Knowledge of purposes of summative assessment	2	2	2	3	3
3. Knowledge of what to assess	1	2	2	3	4

4. Knowledge of assessment strategies and task design	1	2	2	3	4
5. Knowledge of assessment interpretation	1	2	2	3	3
6. NCEA assessment	1	2	2	3	3
7. Preparing students for standards-based assessment	0	2	2	3	3
8. Using summative assessment formatively	2	3	3	3	3
9. Understanding assessment consequences	1	1	2	2	2
10. Fairness	1	1	1	2	3

Mary	Int 1	Int 2	Int 3	Int 4	Int 5
1. Ability to describe assessment	1	2	2	3	3
2. Knowledge of purposes of summative assessment	1	2	2	3	3
3. Knowledge of what to assess	1	2	2	2	3
4. Knowledge of assessment strategies and task design	1	2	2	2	3
5. Knowledge of assessment interpretation	1	2	2	2	3
6. NCEA assessment	1	1	2	2	3
7. Preparing students for standards-based assessment	0	2	2	3	3
8. Using summative assessment formatively	1	2	2	2	3
9. Understanding assessment consequences	1	1	2	2	2
10. Fairness	1	1	1	2	3

Susie	Int 1	Int 2	Int 3	Int 4	Int 5
1. Ability to describe assessment	1	2	2	3	3
2. Knowledge of purposes of summative assessment	1	3	3	3	3
3. Knowledge of what to assess	2	2	2	3	3
4. Knowledge of assessment strategies and task design	1	3	3	4	4
5. Knowledge of assessment interpretation	2	2	2	3	3
6. NCEA assessment	1	2	2	3	4
7. Preparing students for standards-based assessment	0	3	3	3	3
8. Using summative assessment formatively	2	3	3	3	3
9. Understanding assessment consequences	1	1	2	2	2
10. Fairness	1	2	2	2	3

Ryan	Int 1	Int 2	Int 3	Int 4	Int 5
1. Ability to describe assessment	1	2	2	2	2
2. Knowledge of purposes of summative assessment	1	1	2	2	3
3. Knowledge of what to assess	1	1	1	2	2
4. Knowledge of assessment strategies and task design	1	1	1	2	2
5. Knowledge of assessment interpretation	2	2	2	2	2
6. NCEA assessment	1	1	1	2	2
7. Preparing students for standards-based assessment	1	1	2	2	3
8. Using summative assessment formatively	2	2	2	2	2
9. Understanding assessment consequences	2	2	2	2	3
10. Fairness	1	1	1	2	2

Wiremu	Int 1	Int 2	Int 3	Int 4	Int 5
1. Ability to describe assessment	1	3	3	3	3
2. Knowledge of purposes of summative assessment	2	3	3	3	3
3. Knowledge of what to assess	1	2	2	3	3

4. Knowledge of assessment strategies and task design	0	2	3	3	3
5. Knowledge of assessment interpretation	1	2	2	3	3
6. NCEA assessment	0	2	2	3	3
7. Preparing students for standards-based assessment	0	2	2	3	3
8. Using summative assessment formatively	2	3	3	3	3
9. Understanding assessment consequences	2	2	2	2	3
10. Fairness	2	2	2	3	3