Beginning Chinese-Heritage Language Teachers' Conceptions and Practices of Assessment

Jiani Yun

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Education, the University of Auckland, 2020

Abstract

Teachers' conceptions of assessment are powerful in influencing the quality of their teaching practices. In the New Zealand context, there have been studies into both preservice and in-service teachers' conceptions of assessment. However, less research has investigated how beginning teachers' conceptions of assessment change from their completion of initial teacher education programmes until they are fully registered. Moreover, research on beginning language teachers' conceptions of assessment has been rarely concerned with those who come from different cultural backgrounds. To address this gap, this thesis investigated how beginning Chinese-heritage language teachers' assessment conceptions and practices as they began to teach the Chinese language in New Zealand secondary schools.

The study consisted of two phases: Phase 1 was a focus group interview with seven preservice Chinese-heritage language teachers who were at the completion of a 1-year initial teacher education programme. The purpose of the interview was to gain understandings of these beginning teachers' conceptions of assessment before they began teaching in secondary schools. Phase 2 was a longitudinal case study investigating three beginning Chinese-heritage language teachers' conceptions and practices of assessment over 2 years as they worked towards fully registered teacher status. In this phase, data were collected through interviews, classroom observations, reflective conversations and documents. A reflexive thematic analysis approach was employed to analyse the data regarding conceptions while a conceptual framework constructed from existing literature guided analysis of the practice data.

The findings demonstrated that these beginning Chinese-heritage language teachers shifted their negative attitudes and deepened understandings about assessment during their 1-year initial teacher education programme. These changes, however, were found to be both helpful and fragile. As the three beginning teachers began to teach in their specific secondary school context, factors of personal confusion about assessment, limited support from school communities (microlevel) and Chinese-heritage cultural influences (macrolevel) constrained beginning Chinese-heritage language teachers from learning about assessment continuously. The findings also highlighted the interactions between the beginning Chinese-heritage language teachers' prior assessment beliefs and their current assessment experiences in the New Zealand context, showing their active attempts to adjust their conceptions and practices to meet the needs of students and schools. All these findings were oriented to continuous support from within and across schools.

This study contributes a new conceptual framework for understanding assessment practice and uses it to describe how beginning Chinese-heritage teachers' assessment conceptions and practices are challenged and developed. It makes recommendations to initial teacher education programmes, secondary school leaders and policymakers, and professional development programmes.

Acknowledgements

Dedication

In memory of my father who has always been in my heart,

I love you.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Introduction

This thesis is an investigation of beginning Chinese-heritage language teachers' assessment conceptions and practices as they completed their initial teacher education (ITE) programme and then over their first 2 years of teaching. Chapter 1 introduces the study, provides background about assessment, the New Zealand assessment context, beginning teachers' professional development and the Chinese assessment context. A rationale for studying beginning Chinese language teachers' assessment conceptions and practices is outlined, and, based upon this, the significance of this study is presented. The chapter concludes with an overview of the thesis.

Assessment, Assessment Literacy, and Assessment Capability

Assessment plays a critical role in ensuring the quality of teaching and learning (Black & Wiliam, 2018; Crooks, 1988; Edwards, 2017; Harlen, 2013). While education systems widely use assessment for summative purposes, including administration, qualifications, and monitoring teaching quality, research has increasingly explored teachers' classroom assessment practices in the last 2 decades in order to find ways to contribute to their improvement. Teachers devote a high proportion of time to assessing student-learning performance and their teaching effectiveness during daily teaching, and having effective assessment practices can also motivate student learning, improve their achievement and cultivate students' capabilities of assessment and learning (Black & Wiliam, 1998; Booth et al., 2014; Crooks, 1988; Edwards, 2017).

Extensive research has explored teachers' assessment literacy, as assessment-literate teachers can enact high-quality assessment and help students with learning outcomes (Edwards, 2017; Fulcher, 2012; Popham, 2009; Stiggins, 1991; Y. Xu & Brown, 2016). In 1991, Stiggins (1991) described assessment-literate teachers as those who "have basic understandings of high- and low-quality assessment and can apply that knowledge to various measures of student achievement" (p. 535). Assessment-literate teachers know what and how to collect, interpret, and report assessment evidence with reliability and validity, and they are able to distinguish sound from unsound assessments. DeLuca and Johnson (2017) describe teacher assessment literacy as a "fundamental competence" to "integrate assessments throughout instruction to support, monitor and report on student learning, and to use summative forms of assessment

to document and demonstrate achievement of educational standards" (p. 121). DeLuca and Bellara (2013) explain that accountability is still dominant in the assessment landscape. Teachers are required to adopt standard-based teaching approaches and student-centred pedagogies to assist students in succeeding in exams and developing learning and assessment capabilities.

Assessment capability, the same term as assessment literacy, has been adopted in the New Zealand assessment context. Booth et al. (2014) distinguish the term *assessment literacy* from *assessment capability*, arguing that assessment literacy is used to describe teachers' appropriation of assessment in practices where "its usage does not always refer to student agency" (p. 139). However, as they explain, developing students' assessment capability through enabling student agency in their assessment is at the heart of New Zealand assessment practices. Therefore, assessment-capable teachers' practices are aiming to "encourage students to feel deeply accountable for their own progress and support [students] to become motivated, effective, self-regulated learners" (p. 140). Booth et al. (2014) provide an understanding that an assessment-capable teacher,

> … helps students to understand what constitutes quality.
> … helps students develop the metacognitive skills to evaluate their work, and
> … helps students learn strategies to modify their own work. (p. 141)

While there have been widespread calls for assessment-capable/-literate teachers, research has found teachers, beginning teachers in particular, are challenged to develop their assessment capabilities in teacher education programmes (DeLuca & Johnson, 2017; DeLuca & Klinger, 2010; Maclellan, 2004). DeLuca and Johnson (2017) argue that beginning teachers are particularly underprepared for developing assessment capabilities, given that assessment education is often overlooked during teacher education programmes. As they point out, factors such as the short and fragmented model of teacher education programmes, and instructors' assessment capabilities, limit the effectiveness of preparation of teachers' assessment capabilities.

As beginning teachers begin to teach in schools, developing assessment capabilities still seems to be challenging (Absolum et al.,2009; Booth et al., 2014). In New Zealand, the Ministry of Education initiated a teacher development programme called Assess to Learn in primary schools, in which teachers were assisted in developing their assessment capability and practices. However, Absolum et al. (2009) found that only limited numbers of in-service

teachers demonstrated their high-level assessment capabilities in the Assess to Learn development programme. Booth (2020) argues that developing assessment capability is likely to be challenging, given that teachers, school leaders, and policymakers are required to make changes to ensure that cultivating students' assessment capability is the centre of the assessment culture. Schools and policymakers need to make changes to understand and support teachers in taking the initiative in assessment practices and developing their assessment capabilities. Teachers need to shift their roles from controlling learning for outcomes to implementing a student-centred approach.

Therefore, assessment capability may, indeed, be even more challenging for beginning secondary school teachers, faced as they are with preparing students for qualification assessments in the last 3 years of schooling.

Assessment in the New Zealand Secondary Schooling Context

At both primary and secondary levels, New Zealand teachers and schools have the flexibility to teach what they deem best for students (Edwards, 2017; Nusche et al., 2012) guided by the *New Zealand Curriculum* (Ministry of Education, 2007). Within this national curriculum, a high priority is placed upon cultivating student assessment capability, and thus teachers are encouraged to use assessment for both formative and summative purposes of advancing teaching and learning (Edwards, 2017). However, studies have shown that teachers are facing more difficulties implementing formative assessment at the secondary school level due to the pressure from qualification examinations (Edwards, 2017; Yates & Johnson, 2018)

In the New Zealand context, most students start their secondary schooling from Year 9 (age 12 to 13) and stay in secondary schools until Year 13 (some students may leave school before Year 13 given that attending school for Years 12 and 13 is not compulsory). Crooks (2002) compares the differences between assessment in primary and secondary schools in the New Zealand context and points out that although primary school students' work is documented for their end-of-year reports, teachers' assessment activities are low-stakes, as these end-of-year reports are not be used for selection purposes. In contrast, secondary school teachers need to mark/grade students' work to produce end-of-term or end-of-year grades or marks. Different stakeholders, such as students, teachers, parents, and schools, are likely to take these grades into consideration when students choose subjects at senior levels.

Three qualifications programmes are operating in New Zealand secondary schools: National Certificate of Education Achievement (NCEA), Cambridge Assessment International

Education (CAIE), or International Baccalaureate (IB). Qualifications from these programmes can be used for applying for entry to universities and in seeking future employment. Examinations in these qualification programmes are standards-based, and therefore, students' success depends on their actual level of knowledge and capabilities.

While CAIE and IB programmes are developed overseas and have been adopted in countries around the world for decades, NCEA is the most common qualification programme in New Zealand; it was introduced by the New Zealand Qualification Authority (NZQA) in 2002 and has gained recognition across countries in recent years. NCEA qualification programmes are guided by the *New Zealand Curriculum*, which was issued in 2007. The *New Zealand Curriculum* specifies eight learning areas: English, the arts, health and physical education, mathematics and statistics, science, social science, technology, and learning languages (Ministry of Education, 2007). These areas aim to develop students' key competencies, such as using language and managing themselves, assisting them to become lifelong learners (Ministry of Education, 2007). The NCEA qualification programme gradually replaced previous qualifications, such as University Entrance and University Bursary qualifications and is now widely adopted in New Zealand secondary schools.

NCEA includes three levels of qualification (Level 1, Level 2 and Level 3) for students from Years 11 to 13, and each level includes a different number of credits that students need to earn to move to the next level (Level 1 is 80 credits, Levels 2 and 3 are 60 credits). While eight learning areas are compulsory for students from Year 1 to Year 10, Year 11 and 12 students need to take six subjects to gain enough credits in Levels 1 and 2, and Year 13 students need to take five subjects in Level 3. For secondary school students, while there is the possibility of making some changes in Years 12 and 13, most subjects have to be decided by Year 11. This is because the Level 2 results that Year 12 students obtain are often used as evidence for entry to universities and future employment (NZQA, 2010). Senior students (Year 12 and 13) carry on learning selected subjects for the final qualification examinations. Therefore, Year 10 students have started to explore suitable subjects in which they are likely to succeed in future qualification examinations. During the process of exploring subjects they will carry on learning, numeracy and literacy requirements for university entrance drive lots of students' subject decisions. Furthermore, guidance from teachers and schools in choosing subjects influences students' decisions. Interestingly, while senior students can gain literacy credits by taking optional subjects, such as dance, they cannot get literacy credits from taking language subjects, such as the Chinese language.

In NCEA programmes, students are assessed on their learning as they complete their selected courses at each level. Both unit standards and achievement standards are employed to assess student-learning performances. Unit standards are usually used in vocational courses (Hara, 2011), and include grades of Not Achieved and Achieved. Achievement standards contain four grades: Not Achieved, Achieved, Merit, and Excellence. NCEA assessments are conducted internally and externally of schools. Secondary school teachers are required to design and implement internal assessments at each level during the year and moderate results with other colleagues to keep consistent with the assessment standards. Students are assessed externally of schools (up to three standards) and experts from NZQA mark their performances in exams.

In New Zealand, 13 schools offer IB diploma programmes for Year 12 and 13. Like NCEA, IB programmes also aim to support students to become lifelong learners. IB students are expected to attain qualities such as take challenges in learning and become self-motivated in learning. Students enrolled in the 2-year IB programme need to complete courses from six subject areas: mathematics, science, studies in language and literature, the arts, individual and societies and language acquisition. They need to perform in assessments at a high level in three to four subjects, and at a standard level in the remainder. Like NCEA assessments, the final exams for IB students are held at the end of Year 13, and their performances in assessments are marked internally and externally of schools. While students complete externally assessed coursework under the supervision of teachers, internal assessment tasks are used throughout the learning and teaching process (IB, n.d.).

Alternatively, students in some schools can choose CAIE programmes to gain their entrance to university programmes. CAIE programmes are offered in around 37 New Zealand schools. CAIE programmes aim to cultivate student's higher-order thinking skills, such as problem solving and independent research. Guided by the Cambridge International Curriculum, Cambridge Upper Secondary programmes within CAIE programmes are offered for secondary students. They are interested in taking Cambridge O level qualification or Cambridge International General Certificate of Secondary Education (IGCSE)—a qualification examination recognised broadly worldwide. Year 11 to 13 students who enrol in upper secondary programmes need to choose up to five subject areas, subjects of mathematics and English are compulsory (the subject of physical education is compulsory in some schools). Unlike NCEA and IB, while schools' teachers assess CAIE students at the end of the academic year, all the marking work is conducted by external experts.

Crooks (2011) argues the critical role of teachers in secondary schools. As he mentions, New Zealand secondary teachers are challenged to be aware of the boundary between being an assessor for tests and being a mentor for learning. It is because most secondary teachers are assessors of internal assessments, and at the same time, they teach students to achieve better scores in internal assessments. In other words, teachers' formative assessment practices may be shaped and constrained by the purpose of achieving in summative assessments.

Beginning Teachers' Professional Development in New Zealand

Beginning teachers are usually referred to as those who have just started their teaching career in schools. In the New Zealand context, beginning teachers often refer to those who have completed ITE programmes and are yet to gain their full practising certificate (Teaching Council of Aotearoa New Zealand, n.d.). Beginning teachers are required to undertake a 2-year programme of induction and mentoring in their first teaching position/s, provided by their school. Wong et al. (2005) inform that beginning teachers in the New Zealand context usually receive 20% release time to participate in professional development workshops, regional meetings, and conferences as part of the induction programme. Experienced in-service teachers who have gained full practising certificates are assigned to beginning teachers as mentors. They work with beginning teachers in their first 2 years, assisting the beginning teachers to develop the capabilities to meet the requirements of the full practising certificate and teacher registration.

Wong et al. (2005) reviewed induction programmes in China, New Zealand, Switzerland, Japan, and France. They comment that the success of induction programmes in New Zealand is attributed to the fact that "they are highly structured, focus on professional learning, and emphasize collaboration" (p. 383). Nevertheless, Grudnoff (2012) argues that the opportunities and resources the New Zealand induction programmes provide for beginning teachers are not always fully utilised. In a qualitative study, Grudnoff (2012) investigated 12 beginning New Zealand primary school teachers' experiences of transition from preservice stage to the in-service stage in their first 15 months through individual and focus group interviews. She found that each beginning teacher has particular learning needs, and resources and support from induction workshops may not suit all beginning teachers. Grudnoff (2012) argues that beginning teachers face retention challenges in schools and develop capabilities during teaching practice. As a result, beginning teachers may shift their focus from improving teaching qualities to establishing teacher roles and acquainting themselves with school requirements (Aarts et al., 2020). Furthermore, Langdon (2011) argues that beginning

teachers face challenges in reconciling the tensions between dealing with classroom management issues and supporting student learning. Therefore, understanding beginning teachers' learning needs and providing ongoing learning opportunities is critical to ensure the success of beginning teachers' professional development.

Chinese-Heritage Language Teachers and the Chinese Assessment Context

The research interest of the present study was to explore beginning Chinese-heritage language teachers' professional growth in assessment as they began teaching in New Zealand secondary schools. While extensive studies distinguish language teachers according to how they acquired such languages (e.g., native and non-native speakers, first- and second-language speakers), the present study aimed to explore cultural influences from Chinese-heritage cultures on beginning Chinese-heritage language teachers' assessment conceptions and practices. Therefore, the term *Chinese-heritage language teachers* was adopted to describe Chinese language teachers whose conceptions and practices of assessment were likely shaped in the Chinese assessment context.

In China, the mainstream education system includes preschool schooling (for students aged from 2–6), primary schooling (for students aged from 6–12), and secondary schooling, which comprises junior secondary schooling (for students aged from 12–15) and senior secondary schooling (for students aged from 15–18). Since 1986, the Chinese government has introduced and promoted the 9-year compulsory education system, which comprises 6 years of primary school education and 3 years of junior secondary school education. Students who complete the 9-year compulsory education need to take the senior secondary school entrance examination (Zhongkao). Their learning competence in Chinese, mathematics, English, physics, history, politics, chemistry, and physical education (some places may add biology and geography) is assessed in this examination. The ranking of students' grades in this examination will decide if they can enter "key" schools. In China, secondary schools comprise both key schools and "ordinary" schools, key schools have reputations for high teaching quality and enrolment rates, and thus receive more recourses from local governments than ordinary schools.

When students enter senior secondary schools, they are required to take nine subjects from Chinese, mathematics, physics, chemistry, biology, geology, English (or other foreign languages), history, geography, politics, music, fine arts, and physical education in their first year of study. As they move toward the end of the first year, students need to choose subjects

that they would like to continue in the following 2 years. Except for compulsory subjects of Chinese, English, and mathematics, students have options to choose one of two: comprehensive natural science (including biology, chemistry, physics) and comprehensive social sciences (ideology and politics, history, geography) as they move toward Year 2. Toward the end of Year 3, students need to take two types of high-stakes examinations: graduation examinations and Gaokao. Those who pass graduation examinations receive certificates of graduation and are eligible to continue taking Gaokao. In Gaokao, students are assessed on their knowledge and skills in four compulsory subjects and comprehensive social/natural science. While junior secondary school students' grades are compared with those of their peers across the state/province, senior secondary school students' grades are compared with those of their peers across the whole country. Therefore, students are likely to learn to achieve high grades in Gaokao.

Chinese authorities have led educational curriculum reform (NCR) since the 1990s. The concept of formative assessment was imported, as traditional thinking towards assessment has been challenged by assessment innovation worldwide. Contradictions between using "exam-oriented" summative examinations and implementing "learning-oriented" formative assessment stand out in the education landscape. In 2001, the Ministry of Education introduced the Compendium of Curriculum Reform for Basic Education (Experimental), which recommends a learner-centred approach to help students become lifelong learners. As Guan and Meng (2007) conclude, one of the outcomes is that nation—states/provinces—schools' three-layered curricula have been confirmed in the reform. Schools and local governments are responsible for designing their own curricula with guidance from the national curriculum. In this way, as Guan and Meng (2007) argue, schools and local governments have more flexibility to make adjustments to meet students' learning needs. Such curriculum changes are also helpful to cater to local students' learning interests and provide support for "their well-rounded" (Tan, 2016, p. 198) development. Teachers are encouraged to collect more evidence from students' performance in all subjects and use different assessments (such as self-assessment and peer assessment) to support student learning (Tan, 2016).

However, the implementation of reform has encountered stiff resistance from the examination system. As Dello-Iacovo (2009) describes, the "'regular system' has proven remarkably resilient, bouncing back with renewed vigour after each assault" (p. 242). On the one hand, school innovations are still restricted by school accountability and parents' expectations of academic achievement, even if they have been granted more liberty to design and implement

an assessment to improve student learning. On the other hand, although assessment practice for improving student learning is encouraged in classrooms, teachers are still under the constant stress of the high-stakes university entrance examination Gaokao. In addition, with little instruction and professional training in assessment, teachers struggle to effectively use assessment tools, such as feedback, to cultivate students' learning and assessment capabilities (Song, 2016; Wu, 1994).

Rationale for the Study

As noted earlier, assessment is an integral part of effective teaching and learning (Black & Wiliam, 1998, 2018; DeLuca et al., 2018; Eyers, 2014; M. F. Hill, 2000). Internationally, teachers are expected to use assessment evidence to support both student learning and establish the accountabilities of teachers. However, as described, this plays out differently in different contexts, and even with supportive policies, innovations are not always easy to embed where policies conflict with institutional and societal expectations. To bring summative and formative assessment together in support of learning and achievement, Volante and Fazio (2007) argue that teachers must have a clear picture of the two purposes of assessment to help them achieve such purposes.

In the New Zealand context, there has been significant research into the assessment conceptions of both preservice and in-service teachers (Brown, 2003; Edwards, 2017; Eyers, 2014; M. F. Hill, 2011). However, less research has investigated beginning teachers' changes in conceptions and practice during the period between ITE and before they are fully registered. Borg (2006) argues that beginning teachers take time to adapt their existing conceptions and practices into different school cultures, partly because their conceptions may have been sustained over a long period and may be resistant to change. Beginning teachers may also experience difficulties integrating what they have learned from teacher education into their classroom practice, and they may also need to adapt their conceptions and practices to specific school cultures. Furthermore, in the New Zealand context, the conceptions of assessment of novice language teachers who come from distinctively different cultural backgrounds are infrequently discussed (Edwards & Edwards, 2017; Sun, 2010). With these issues in mind and given the increasing numbers of Chinese-heritage language teachers entering the New Zealand system, it is vital that beginning Chinese-heritage language teachers' conceptions of assessment are investigated both at the conclusion of their ITE programme and as they begin teaching. This thesis set out to investigate these conceptions and the practices that followed from them. Therefore, the research question that guided this study was:

What are beginning Chinese-heritage language teachers' conceptions and practices of assessment at the completion of their 1-year initial teacher education programme and as they begin teaching over the first 2 years?

To investigate beginning teachers' conceptions and practices of assessment, the present study included two successive phases. Phase 1 explored preservice Chinese-heritage language teachers' assessment experiences. The overarching question used to shape the Phase 1 study was:

What are preservice Chinese-heritage language teachers' conceptions of assessment at the completion of their 1-year initial teacher education programme?

The specific research questions in Phase 1 were:

1) What are preservice Chinese-heritage language teachers' conceptions of assessment at the completion of their 1-year initial teacher education programme?

2) What are the factors influencing preservice Chinese-heritage language teachers' conceptions of assessment?

3) How well prepared do they feel to put their assessment understandings and skills of assessment into classroom practice?

The Phase 2 study aimed to explore beginning Chinese-heritage language teachers' assessment experiences, and the guiding question was:

What are beginning Chinese-heritage language teachers' assessment conceptions and practices and how do these change as they progress through their first 2 years of teaching?

The specific questions used in Phase 2 were:

1) What are beginning Chinese-heritage language teachers' conceptions of assessment as they move towards full teacher registration?

2) What are beginning Chinese language teachers' practices of assessment as they move towards full teacher registration?

3) What factors influence beginning Chinese-heritage language teachers' assessment conceptions and practices as they begin teaching?

Structure of this Thesis

This thesis is structured in seven chapters. The introduction chapter has introduced the research context, summarised how different contexts might influence beginning Chinese-heritage language teachers' assessment conceptions and practices, and provided the research questions that guided the study.

Chapter 2 reviews the literature in three areas that contribute to answering each of the research questions in the two phases. That is, literature is reviewed with regard to assessment, beginning teacher's conceptions and practices of assessment, and the factors that influence beginning teacher's conceptions and practices of assessment. Furthermore, a conceptual framework is described based on a review of previous assessment conceptual frames. The new conceptual framework is described as it was then used for analysing the practices data in this thesis.

In Chapter 3, the study design and research process are presented. The study's methodology and methods are introduced, explaining why this study fits within the interpretive paradigm and case study approach. Then, a detailed description of the study design, participants, instruments, and the procedure of data collection in each phase is explained. Next, the reflexive thematic approach to data analysis is introduced, and the data-analysis process is explained. Finally, trustworthiness and ethical considerations for the current study are presented.

Chapter 4 presents the findings of the investigation into the preservice Chinese-heritage language teachers' assessment conceptions. Their assessment conceptions at the exit from the 1-year initial teacher education programme were gathered, and changes in assessment conceptions and practice during the programme are presented. Furthermore, this chapter describes how well the beginning Chinese-heritage language teachers feel prepared to use assessment in secondary schools.

Chapter 5 shifts to answer the research question regarding the beginning Chinese-heritage language teachers' conceptions of assessment as they began to teach in New Zealand secondary schools. Changes in their conceptions and reported practices of assessment are collected and presented based on the timeline of four stages as they moved toward full teacher registration.

In Chapter 6, data regarding the beginning Chinese-heritage language teachers' observed assessment practice is presented for each case. Their assessment practices are presented from

three aspects: their assessment purposes, processes, and forms, as described in the conceptual framework for this study.

Chapter 7 discusses the findings of both phases regarding relevant literature. Factors that influence beginning Chinese-heritage language teachers' assessment learning are discussed from personal, micro-, and macrolevels. Finally, conclusions, implications, limitations, and recommendations for researchers, beginning Chinese-heritage language teachers, secondary school principals, and decision makers are addressed.

Chapter 2 Literature Review

Introduction

Chapter 1 presented an overview of the research background and the context of this study. In Chapter 2, literature centred around assessment and development of preservice and beginning teachers' assessment conceptions and practices is reviewed. Increasing numbers of empirical studies focus on preservice and in-service teachers' development of assessment conceptions and capabilities, while less is known about how beginning teachers change or develop their conceptions and practices of assessment in their early teaching careers. Furthermore, although previous literature explores factors contributing to changes in teachers' conceptions and practices of assessment, limited research, such as Sun (2010), explores the influences Chinese heritage might have on teachers' conceptions and practices of assessment.

In this chapter, literature regarding assessment is reviewed in four sections. Initially, theoretical ideas about assessment purposes are addressed. Teachers' conceptions and practices of assessment are examined next, including studies that have investigated preservice and beginning teachers' conceptions and practices of assessment. Third, the influences of personal, school, policy and culture on assessment are presented. Finally, a conceptual framework for this study is described based on a review of previous frameworks of teachers' assessment practices.

Theoretical Understanding of Assessment Purposes

Assessment is considered essential to enhancing and recording the outcomes of teaching and learning (Black & Wiliam, 1998; Eyers, 2014; Maclellan, 2001; McMillan, 2008; Taras, 2005). Previous literature has explored assessment activities based on different purposes. In the UK context, the Task Group on Assessment and Testing for England and Wales (1988), summarised assessment activities as four types: formative assessment, diagnostic assessment, summative assessment, and evaluative assessment. The Assessment Reform Group (2002) provided two distinct terms for assessment purposes: assessment of learning and assessment for learning. Meanwhile, other terms such as dynamic assessment and non-dynamic assessment (or static assessment) have also been popular and used in the U.S. context (e.g., Poehner, 2008).

Although various terms are used to describe assessment activities to suggest different assessment purposes, such as diagnostic and evaluative assessment, formal and informal

assessment, *formative assessment* and *summative assessment* are most often used in literature to indicate two fundamental assessment purposes—assessment to inform learning and assessment to check on/sum up learning (Black &Wiliam, 1998, 2018; Cowie & Bell, 1999; Eyers, 2014; Gardner & Harlen, 2010; Harlen, 2007). Daugherty and Ecclestone (2006) suggest that instead of the more technical formative and summative assessment, the terms *assessment for learning* and *assessment of learning* are more accessible. However, as Black and Wiliam (2018) argue, the terms assessment for learning and assessment of learning can barely support readers to understand the important role that peers need to play in supporting learning during assessment activities, and thus they "serve to obscure rather that illuminate the important processes" (p. 3).

Based on the need for deep understanding of assessment purposes, the current study adopted Black and Wiliam's (2018) explanations, using formative and summative assessment as the two terms for describing and understanding the two fundamental purposes of beginning Chinese-heritage language teachers' assessment practices. In brief, these mean the summative accountability purpose (of schools, teachers, and students), the formative purpose of using assessment to improve students' learning and teachers' teaching, and other purposes such as classroom management. Each of these purposes and their effects are discussed next in this review.

Summative Assessment

Literature has provided definitions of summative assessment from different perspectives. From the perspective of summative assessment processes, Crooks (2011) describes summative assessment process "involves judging, describing and reporting student educational outcomes (such as their knowledge, skills and attitudes) at a particular point of time" (p. 71). Teachers' judgements toward student educational outcomes at one point of time are made for comparisons with other time points during summative assessment processes. From the perspective of summative assessment purposes, Gardner and Harlen (2010) identify summative assessment by its two major uses: "Accountability—evaluation of teachers, schools, local authorities against targets; (and) Monitoring—students' average achievements within and across schools in particular areas, and across a whole system for year-on-year comparison" (p. 16). They further point out the potential risks when teachers implement assessment activities for these two purposes. First, teachers who only hold beliefs about accountability purposes of assessment are unlikely to advance student learning through assessment activities. Their classroom practices may demotivate student learning when their

focus is on student-learning outcomes rather than advancing learning. With regards to the monitoring purpose of summative assessments, tests or examination results may not reflect the changes of individual student's learning performances, given that teachers may use summative assessment activities to monitor performances of student groups rather than individuals. Thus, refining national standards through students' test results is unlikely to be achieved.

In the U.S. context, Garrison and Ehringhaus (2007) highlight the "control" characteristics of summative assessment. They list forms of summative assessment includes: state assessments, district benchmarks or interim assessments, end-of-unit or chapter exams, end-of-term or semester exams, and scores used for accountability of schools and students (report card grades). They point out that summative assessment has been successfully used for evaluation and administration purposes in schools and institutions and, for many decades, it was understood as the major way to assess student performance. The results of this type of assessment provide powerful evidence to improve the effectiveness of learning programmes, school improvement, and alignment with curricula. Furthermore, Garrison and Ehringhaus comment that summative assessment results "happen too far down the learning path to provide information at the classroom level and to make instructional adjustments and interventions during the learning process" (p. 2).

Many studies have explored the effects of summative assessment on students, teachers and schools in the New Zealand context (e.g., Brown et al., 2009; Crooks, 2011; Yates & Johnston, 2018). In this context, Crooks (2011) reviews innovations of the assessment system that New Zealand has made from the previous high-stakes examination system to a comparatively lower stakes system of the NCEA in the last 2 decades. He argues that the pressure from nationwide NCEA exams still influences secondary students' decisions about what subjects they will take. Students may choose subjects that they are more likely to succeed in, in NCEA exams, and schools which are aiming to raise their success percentage may guide students to choose subjects or standards within subjects they are more likely to pass. Furthermore, with the increasing numbers of students taking international examinations such as IB and CAIE for qualification purposes (see more details about systems of IB and CAIE in Chapter 1, p. 4), M. F. Hill et al. (2018) raise the concern about the high pressure of accountability from international high-stakes examinations on students, teachers and schools. They argue that comparisons of student achievements in international high-stakes examinations drive teachers

and students to work for tests and may further strengthen the tension between national examinations and teachers' implementation of formative assessment.

Despite the fact that summative assessment could be used to control learning and monitor teaching qualities, conceiving it as the opposite of formative assessment hinders making use of summative assessment data. Harlen (2013) argues that summative assessment is critical for students, teachers and schools to trace student learning. Schools evaluate teaching performances and make improvements based on summative assessment data and corresponding contextual factors. Also, students can gain ownership of their learning through summative assessments. However, as Harlen asserts "The bad reputation of summative assessment arises from inappropriate use of data which do not fully reflect the goals of learning" (p. 20). The methods of collecting summative assessment information may influence how people understand summative assessment in a comprehensive way, and conventional methods of summative assessment may not be able to reflect various learning goals (Harlen, 2013).

Formative Assessment

Formative assessment has been advocated increasingly with a focus on improving student-learning processes (Black & Wiliam, 1998; Cowie & Bell, 1999; Crooks, 1988; Gardner, 2006; Harlen, 2007). Black and Wiliam's (1998) seminal study reviewed previous literature about formative assessment, defining formative assessment as "all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged" (p. 10). In a longitudinal study regarding teachers' formative assessment practices in six UK secondary schools, Black et al. (2003) confirmed the significance of implementing formative assessment in real classroom settings and concluded three benefits of using formative assessment as being: a focus on learning, where the teacher's task is to involve students in learning and to help them build and reconstruct their knowledge; a change of role, referring to teachers' role change from knowledge-delivery authority to knowledge-sharing partner; and motivation and self-esteem, which includes the outstanding feature of formative assessment—that assessment could provide feedback to students within their day-to-day classroom life.

Black et al. (2003) highlight that teachers' formative assessment practices could effectively build up student confidence as well as motivate them to be self-learners. Similarly, Harlen (2013) points out that students can effectively improve their capabilities of self-assessment

and responsibilities of learning when teachers adjust learning goals to match students' existing ideas and skills and share their ownership of assessment (data collection, interpretation and feedback) with students. To inform and enable teachers' formative assessment practices in classrooms, M. Thompson and Wiliam (2007) propose five formative assessment strategies: (1) clarifying learning intentions and sharing criteria for success; (2) engineering effective classroom discussions, questions, and learning tasks that elicit evidence of learning; (3) providing feedback that moves learners forward; (4) activating students as the owners of their own learning; (5) activating students as instructional resources for one another (see details in Figure 2.1).

Figure 2.1

The Five Key Strategies in Teacher Assessment

	Where the learner is going	Where the learner is right now	How to get there
Teacher	Clarifying learning intentions and criteria for success	Engineering effective classroom discussions, questions, and learning tasks that elicit evidence of learning	Providing feedback that moves learners forward
Peer	Understanding learning intentions and criteria for success	Activating students as instructional resources for one another	
Learner	Understanding learning intentions and criteria for success	Activating students as the owners of their own learning	

Note. Adapted from M. Thompson and Wiliam (2007, p. 7)

The "big idea" behind the five formative assessment strategies is to help teachers enable formative assessment in classrooms and to adapt teaching to meet student-learning needs (M. Thompson & Wiliam, 2007). However, as they suggest, teachers may not be able to make use of these assessment strategies in classrooms when they have limited understanding and knowledge of formative assessment.

In the New Zealand context, both teachers and students are encouraged to use formative assessment for learning, as the "heart" of the New Zealand assessment culture is to promote student learning and cultivate their assessment capabilities (Booth et al., 2014). As explained in the Ministry of Education (2010) Position Paper, New Zealand has an assessment focus that is different from other countries. Teachers' professional judgement, underpinned by formative assessment, is the focus of all assessment practices and influences how standard-based examinations are carried out. Teachers are encouraged to implement formative assessments in the classroom, as formative assessment is "the process of seeking and interpreting evidence for use by learners and their teachers to decide where learners are in

their learning, where they need to go and how best to get there" (Assessment Reform Group, 2002, p. 2).

However, formative assessment, as Yorke (2003) points out, is a concept "more complex than it might appear at first sight" (p. 478). Formative assessment comprises two forms: planned and unplanned. Planned formative assessment is usually considered formative assessment and includes three essential steps of formative assessment, as Wiliam (2011) summarises:

> 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 what was elicited. (p. 6)

Cowie and Bell (1999) suggest that planned formative assessment occurs through classroom activities arranged by teachers. Furthermore, students are able to gain information when teachers make assessment decisions and take actions. When students act on assessment information, teachers can further collect student information. The reciprocal nature of planned formative assessment encourages teachers and students to learn collaboratively.

With regard to unplanned forms of formative assessment, previous studies identify this type as interactive formative assessment (Cowie & Bell, 1999), informal formative assessment (Ruiz-Primo & Furtak, 2006), incidental formative assessment (Ellis, 2006) or spontaneous formative assessment (Black & Wiliam, 2009; Clark, 2012; Purpura & Turner, 2015). Regardless of terminology, this type of formative assessment is more constructivist in orientation (Carless, 2007), and defined as teachers and students' interactions, happening "over a moment" and imbedded in teaching and learning activities (Clark, 2012; Cowie & Bell, 1999; Ruiz-Primo & Furtak, 2006). Interactions between teachers and students or between student peers during unplanned formative assessment, students to stretch over the "zone of proximal development"[1] (Clark, 2012).

Extensive literature has explored teachers' formative assessment practices and showed that teachers are challenged in understanding and implementing both planned and unplanned

---

[1] Zone of proximal development: Vysgotsky (1978) identifies ZPD as the gap between "actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (p. 86).

formative assessment (Black & Wiliam, 2018; Cowie & Bell, 1999; Harlen, 2013; M. F. Hill, 2000; Taras, 2007; M. Thompson & Wiliam, 2007). First, teachers who conceive formative assessment as a series of classroom techniques may be at risk of overlooking the principles of formative assessment (Cooper & Cowie, 2009; Lee & Wiliam, 2005; Marshall & Drummond, 2006; M. Thompson & Wiliam, 2007). For example, Marshall and Drummond (2006) describe the principles and techniques of formative assessment as the "spirit" and "letter" of formative assessment. Teachers who understand that the spirit of formative assessment is to improve learner autonomy make better use of information emerging from classroom activities. Similarly, M. Thompson and Wiliam (2007) mention that understanding the techniques of formative assessment—the letter ("know how")—is necessary but not enough for teachers to implement formative assessment activities in classrooms. Understanding the principles of formative assessment ("know why") could "empower them to make implementation decisions that enhance, rather than detract from, the theory of action" (p. 7). In contrast, teachers who lack understanding of the principles of formative assessment may tend to control classrooms and implement routinised assessment activities (Marshall & Drummond, 2006; M. Thompson & Wiliam, 2007). Furthermore, teachers may be challenged to understand unplanned formative assessment. They may conceive unplanned formative assessment as part of good teaching (Carless, 2007, Cowie & Bell, 1999). As Cowie and Bell (1999) stress, teachers' degrees of awareness of unplanned formative assessment impact on their practice of making use of unplanned formative assessment activities. In addition, teachers' formative assessment practices are also connected with different social contexts, as Davison and Leung (2009) comment: "it is extremely difficult to sustain any significant teacher-based formative assessment practices in most traditional examination-dominated cultures" (p. 398).

Integrating Formative and Summative Assessment into Teachers' Teaching Practice

Despite formative assessment being widely recognised as learning oriented, while summative assessment is linked with the purposes of accountability, formative and summative assessment should not be considered as two distinct, or even opposite, concepts. Instead, both formative and summative assessment could be used mutually to advance student-learning autonomy (Black et al., 2003; Black & Wiliam, 2018; Deluca et al., 2012; Gardner & Harlen, 2010). Black and Wiliam (2018) explain the relationship between formative and summative assessment by clarifying the inference nature of assessment:

> *Assessment*—a procedure for making inferences. The purpose of the assessment may be broken up into a collection of many different aspects—some of these inferences

will be about the stage of learning that a student has reached, and other inferences will

be related to the courses of action that would best advance a student's learning—but

it is not obvious that this collection either could be, or should be, neatly divided into

two distinct groups. (p. 4)

In their view, the distinctions between formative and summative assessment become evident at the stage that teachers make use of collected assessment evidence. Therefore, the distinction between these two assessment purposes may not be necessary as long as assessment information is used for learning.

Furthermore, Black et al. (2003) argue that it is unrealistic for teachers to separate summative assessment, such as tests and examinations, from the rest of their assessment practice. They investigated UK teachers' conceptions and practice of assessment and found that teachers had difficulties taking up formative assessment when conceiving formative and summative assessment as distinct processes. Under the pressure of exams and tests, formative assessment data were used for summative assessment purposes, "where teachers had control over the setting and the marking" (p. 12). Black and Wiliam (2018) further explain that teachers tend to adopt any approach to solve assessment problems and advance student learning, showing teachers act on assessment evidence in an integrated way rather than distinguish summative assessment from formative assessment purposes first. Similarly, Taras (2007) suggests that a deliberate separation of formative and summative assessment purposes can hinder the implementation of formative assessment, given that the definition and process of formative assessment remain somewhat unclear in the literature:

The practical and logistical problems for tutors of not having watertight definitions of

the terminology and the processes of assessment are enormous. When tutors perceive

that (or are led to believe that) formative assessment is something different from

summative assessment, rather than a logical (often short) next step, few will jump in

willingly to double their workload. (p. 370)

Beyond the discussion about the relationship between formative and summative assessment, integrating assessment with pedagogy or teaching is also critical to the effectiveness of assessment (Black & Wiliam, 2009, 2018; Chan, 2019; Harlen, 2013). Assessment used to be considered a "stop sign" of the education process and thus with limited connections to pedagogy and teaching (Harlen, 2013). However, as Tung (2019) suggests, new assessment and pedagogy thinking and practice are expected to meet student-learning needs with the shift from behaviourist to constructivist approaches to teaching and learning. Black and Wiliam

(2018) propose a framework for considering assessment processes within the context of pedagogy and teaching (see Figure 2.2).

Figure 2.2

Model for Assessment in Relation to Pedagogy
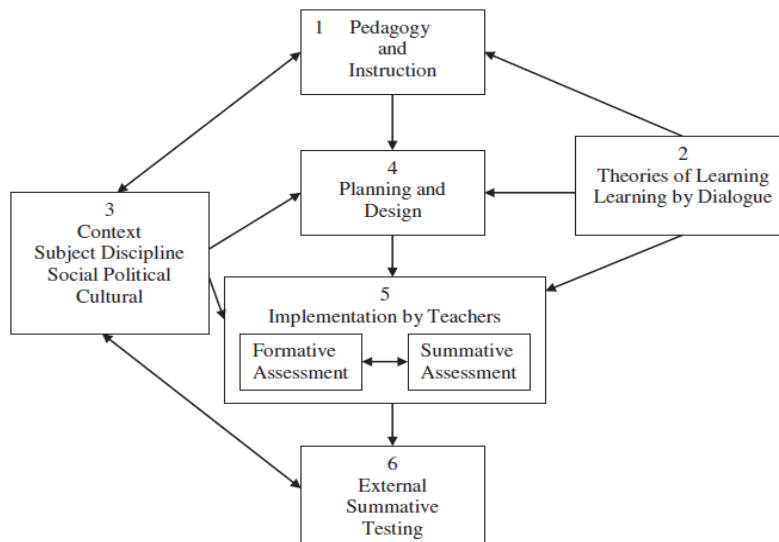
Note. Adapted from Black and Wiliam (2018, p. 556)

They assert that assessment plays a critical role in effective education due to the gap between teachers' instructions and actual knowledge that students have learned. The differences between formative and summative assessment become evident when teachers make decisions and take actions with assessment information (e.g., Carless, 2007; C. Turner & Purpura, 2015). Based on this understanding, the following section presents other purposes of assessment.

Other Purposes of Assessment

Assessment data is also used for other purposes (Brookhart, 2001; Cheng et al., 2004; K. Hill, 2017; K. Hill & McNamara, 2012; McNamara, K. Hill & May, 2002; Rodriguez, 2019). Brookhart (2001) indicates that in the teacher-created classroom assessment environment, their assessment practices integrate with their instructional and classroom management practices. K. Hill and McNamara (2011) synthesise research and point out that teachers use assessment data for learning, teaching, reporting, management and socialisation. In their investigation of primary and secondary Indonesian language teachers' assessment conceptions and practices of assessment in the Australian context, they found that teachers use assessment information for classroom management, encouraging students to work hard.

Furthermore, teachers were found using assessment information to socialise students into new teaching and assessment cultures.

Interestingly, teachers' uses of assessment information for management purposes seem to be related to contextual factors. In a comparison study of English as foreign language (EFL) teachers' assessment conceptions in Canada, Hong Kong and Beijing, Cheng et al. (2004) used questionnaires to find that 94.8% of English teachers in Beijing prefer to use assessment data to encourage students to work hard, while their counterparts in Canada and Hong Kong use assessment for such purposes less, 65% and 70% respectively They assert that while EFL teachers in Hong Kong and Canada use assessments for student-centered purposes, teachers in Beijing put emphasis on instructional purposes; the difference is explained by the different influences of external factors such as the environment of learning and teaching as well as the role of external tests.

Cheng et al. (2004) explore the differences between EFL teachers within different contexts, showing contextual factors impact teachers' take up of assessment for classroom management purposes.

Assessment Research in Second-Language Education

Along with extensive research focused on assessment in general education, theorisation regarding assessment in the second-language field has also developed over 20 years (Carless, 2007; Davison, 2004, 2019; Dickins, 2001, 2006; Leung, 2004), with a particular interest in exploring the integration of both formative and summative assessment to support learning and teaching (Carless, 2012; Davison, 2019; Purpura & Turner, 2018; Zeng et al., 2018) and types of formative assessment; (Dickins, 2001, 2006; Davison & Leung, 2009; K. Hill & McNamara, 2012; Leung & Mohan, 2004; Purpura & Turner, 2018). Research regarding types of formative assessment and the integration of formative and summative assessment in the second-language assessment field are introduced in the following two subsections.

Learning-Oriented Assessment in the Second-Language Assessment Field. Alonzo (2016) summarises three theoretical development waves in the assessment landscape. In the first development wave, debates were around the distinction between formative and summative assessment. Discussions around providing effective feedback in formative assessment were highlighted in the second development wave. More recently, the third development wave has underlined the change in assessment research from understanding formative and summative assessment as two distinctive purposes to integrating both purposes of assessment.

In response to such innovation, in the second-language assessment field, several researchers explore teachers' practices of formative and summative assessment in language classrooms. Carless (2007) attempts to integrate both formative and summative assessment to advance learning in tertiary classrooms by proposing a framework of learning-oriented assessment (LOA) (see LOA framework below in Theoretical Frameworks of Assessment section for more detail). As he points out, formative and summative assessment purposes are not against each other, both assessment purposes can be realised when assessment activities promote sound student learning; students can be encouraged to be responsible for their own and/or peer learning using results from summative assessments; and feedback from teachers and students can also be used as feedforward for future learning and teaching.

Furthermore, C. Turner and Purpura (2015) propose a working framework of LOA based on Carless's work (see Figure 2.3). They assert that classroom assessment involves questions regarding assessment purposes (such as formative and summative purposes), agents (such as teachers, students and parents etc.) and dimensions (such as classroom context). Therefore, their proposed framework includes seven dimensions to understand and enact assessment from different perspectives: the contextual dimension (i.e., the social, cultural, and political context of learning); the elicitation dimension (i.e., planned and unplanned elicitation); the proficiency dimension (i.e., forms and uses of language elements); learning dimension (i.e., the role of feedback and self-regulation); instruction dimension (i.e., knowledge of the target language and the subject matter); interactional dimension (i.e., the structure of interactions, the assistance after interactions); and affective dimension (i.e., learners' emotions, attitudes and motivations). As they explain, teachers' assessment practices are complex and the deconstruction of LOA into seven dimensions is helpful to understand the complex nature of classroom assessment. In this way, classroom assessment can be understood and used to support learning in an effective manner in second-language classrooms.

Figure 2.3

Working Framework of Learning Oriented Assessment (LOA)
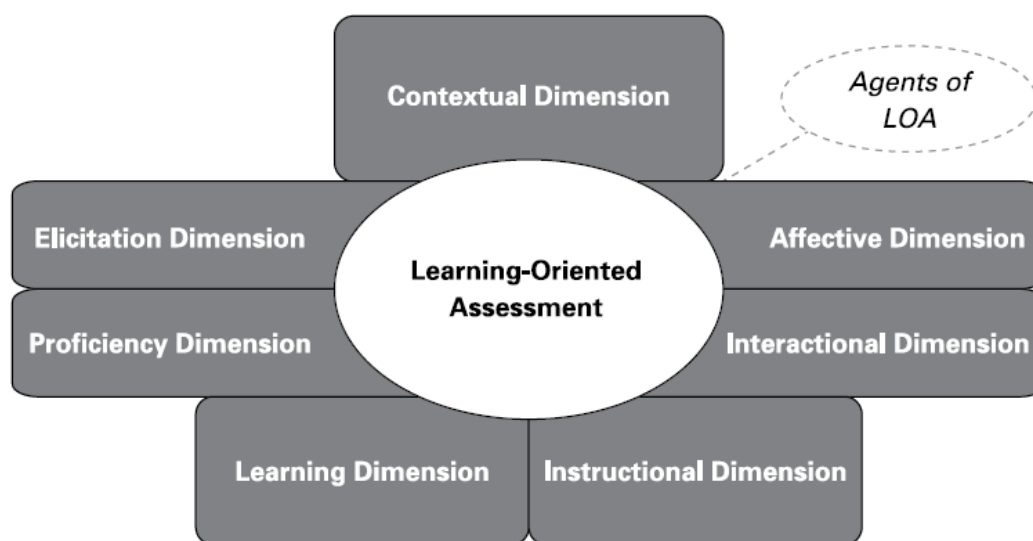
Note. Adapted from C. Turner and Purpura (2015, p. 261)However, as Davison (2019) comments, there is a clear gap between language assessment theories and effective assessment practices. Teachers' practices of formative assessment are constrained by their limited understandings regarding the spirit of formative assessment. C. Turner and Purpura (2015) assert that language teachers' assessment decisions still stem from their intuition and practical experience, given that there is limited research on the structures of unplanned forms of assessment or spontaneous assessment, and the effect of using such assessment to advance second-language learning.

Unplanned Formative Assessment Research in the Second-Language Assessment Field. While formative assessment is promoted in the second-language assessment landscape, limited studies explore different types of formative assessment (Davison, 2019; Davison & Leung, 2009; Dickins, 2001; K. Hill & McNamara, 2012; Purpura & Turner, 2018; C. Turner & Purpura, 2015). Dickins (2001) proposes a framework of classroom assessment processes (see more about this framework in later this chapter) based on an investigation of three English-language teachers' classroom assessment practices. During her observations of language teachers' classroom assessment, participant teachers reflected that they used "assessment opportunities" to support student learning. These assessment opportunities were used by teachers to realise purposes such as tracking student progress against standards or adjusting instructions for individual students. K. Hill and McNamara (2012) broaden Dickins' (2001) notion, providing a definition of assessment opportunity:

Any actions, interactions or artefacts (planned or unplanned, deliberate or unconscious, explicit or embedded) which have the *potential* to provide information on the qualities of a learner's (or group of learners') performance. (p. 397)

In their investigation into second-language teachers in the Australian context, both planned and unplanned/incidental forms of formative assessment were found in classroom assessment opportunities, showing that classroom assessments are not only carried out through teachers' delicate planning and conduction, but also emerge from their intuition. In addition, K. Hill and McNamara propose a comprehensive framework of classroom-based assessment to guide teacher practitioners and researchers to explore the complex nature of language teachers' assessment practices.

From a different perspective, Dickins (2006) and Leung and Mohan (2004) call for more research to inquire into a type of formative assessment that is embedded in instructions. Dickins (2006) conducted a longitudinal qualitative study to observe three language teachers' classroom assessment practices and their students' learning performances over a 12-month period. The findings reveal that many assessments were routinely embedded in teachers' instructions; language teachers used such assessments to recognise the extent to which students had learnt and further adjusted instructions accordingly. As Dickins points out, limited understanding that instruction-embedded assessment is for improvement of teaching or curriculum purposes constrains teachers' ability to engage students in such assessment events; thus the theorisation of different forms of formative assessment plays a critical role in recognising and understanding language teachers' formative assessment practices.

In response to the research gap regarding types and forms of assessment, Purpura and Turner (2018) developed a framework of types of assessment (see Figure 2.6). In their framework, interactive formative assessments or spontaneous assessments are embedded in teaching instructions or conversations. As C. Turner and Purpura (2015) comment regarding language teachers' adoption of interactive formative assessment "the nature and structure of these sequences are important as they may contribute to the effectiveness of feedback (i.e., evaluation and assistance) on learning success" (p. 268).

Davison and Leung (2009) review literature and point out that language teachers' formative practices place a high preference on using the approach of interactions between teachers and students to advance student learning in language classrooms. In other words, interactive formative assessments are frequently undertaken by language teachers to advance student

25

learning (Davison & Leung, 2009; Dickins, 2001, 2006). Language teachers who adopt such an approach are encouraged to interact with students at any time to enhance students' competence in oral interactions by asking clarifying questions. However, this approach is mainly implicit and relies on teachers' intuitions, which raises the question of whether language teachers' conceptions of assessment influence how they conceive and use interactive formative assessment. Furthermore, as Davison and Leung (2009) assert, the interactive formative assessment is not evident in teacher-based assessment models, language teachers may be confused about enabling such practice in classrooms.

Teacher Conceptions and Practices of Assessment

Conceptions, sometimes called beliefs, "put metaphysical beliefs and concepts in a central place, and filter new information before knowledge is acquired" (Pajares, 1992, p. 320). A. G. Thompson (1992) distinguishes teachers' conceptions from teachers' beliefs by referring to teachers' conceptions as complex mental structures including beliefs and concepts. He explains that teachers' conceptions are consciously and subconsciously represented by "their beliefs, concepts, meanings, rules mental images and preferences concerning the discipline of mathematics" (p. 132). In alignment with this understanding, Brown (2002) identifies teachers' conceptions as "an organising framework by which an individual understands, responds to, and interacts with a phenomenon" (p. 2). He later explains that teachers' conceptions of assessment include beliefs; "it may be that one of the most important changes that prospective teachers need to go through involves their conceptions, that is, their understandings, beliefs, and attitudes towards the use of assessment" (Brown & Remesal, 2012, p. 76). More recently, Barnes et al. (2015) provide further explanation of teacher conceptions, stating "teachers' conceptions essentially consist of several concepts, in which knowledge and beliefs are included and understood as teachers' awareness and perceptions" (p. 285). Barnes et al.'s definition of conceptions is helpful in situating an understanding of beginning Chinese-heritage language teachers' knowledge, beliefs and attitudes about assessment.

Nasr et al. (2020) conducted a quantitative study to investigate Iranian English as a second language (ESL) teachers' formative assessment conceptions and practices with a specific focus on class size and gender. A total 384 secondary ESL teachers, ranging in age from 25 to 60, responded to the designed questionnaire. The results show that while both female and male secondary ESL teachers conceived assessment as a tool for monitoring and scaffolding student learning, female teachers, in particular, were found to favour using formative assessments to scaffold student learning more than male teachers. In contrast, there were only

marginal differences in statistics between female and male teachers when they used assessment for monitoring purposes. Furthermore, secondary English-language teachers' responses highlighted that their use of assessment information in the classroom was not influenced by class size. Nasr et al. (2020) explored how factors of gender and class size influence language teachers' assessment conceptions and practices in the Iranian context; nevertheless, factors other than gender and class size that influence language teachers' conceptions of assessment are yet to be explored. Furthermore, their questionnaires only collected secondary English teachers' self-reported assessment practices, and thus the validity of collected data might be compromised. More studies are needed to observe language teachers' assessment practices in classrooms.

In the Turkish context, Büyükkarcı (2014) conducted a quantitative study to investigate primary English-language teachers' conceptions of formative assessment. Through the employment of a mixed-methods approach, 69 in-service primary English teachers from different primary schools completed the designed questionnaire; 10 English-language teachers were interviewed individually. The results showed the discrepancy between primary English-language teachers' beliefs about formative assessment and their self-reported practices. While teachers embraced formative assessment and showed positive attitudes towards using formative assessment, their self-reported classroom practices showed that they used assessment for summative purposes. Büyükkarcı argues that the primary English-language teachers needed to teach more than 20 hours every week, and their classes were crowded (more than 60 students in a class). Heavy workload and crowded class sizes prevented the primary English-language teachers from implementing formative assessment in classrooms. As a result, primary English-language teachers chose to adopt school-designed tests to assess student learning rather than designing and using formative assessment activities. Büyükkarcı's (2014) study explores second-language teachers' conceptions of assessment in the Turkish context and the factors that influence teachers' enactment of formative assessment in practice. More studies are needed to explore second-language teachers' conceptions and practices of assessment in other contexts and in the secondary school sector.

With a different study emphasis, Chen and Brown (2016) explore the interrelationship between teachers' assessment conceptions and their teaching approaches in the Chinese context. A self-report survey of approaches to teaching inventory (ATI) was conducted, and 1,500 sciences, social sciences, and humanities teachers from primary schools, middle schools, and high schools responded to the questionnaire. The results indicate that Chinese teachers

are likely to adopt teaching approaches of knowledge transmission and examination preparation and be student focused, whereas their assessment conceptions highlight the purposes of student development and teaching improvement. Chen and Brown (2016) explain that Chinese teachers who adopt the teaching approach of knowledge transmission are likely to conceive accountability and examination as a powerful method to improve learning and teaching in the examination-oriented context. Such a teaching approach can be used to predict students' performance in examinations and as "an indicator of moral and personal worth" (Chen & Brown, 2016, p. 362). However, the relatively low level of endorsement of the knowledge-transmission approach, revealed in statistics shows that teachers do not desire to use examinations to control student learning. In contrast, teachers who hold beliefs about teaching with a student-centred approach, conceive assessment as positively related to improving student learning and refining teaching for better learning outcomes. Chen and Brown (2016) argue that Chinese teachers' endorsement of both the knowledge-transmission approach and student-centred approach show strong correlations between their conceptions of assessment for accountability and for improvement. Chen and Brown (2016) shed light on how Chinese teachers conceive and take up different purposes of assessment; nevertheless, more research is needed to explore Chinese-heritage language teachers' assessment conceptions in other contexts.

While reviewed studies explore language teachers' conceptions of assessment and Chinese teachers' assessment experience, there is no study regarding Chinese language teachers' assessment conceptions and practices in the New Zealand context. In the assessment context of other subjects, M. F. Hill (2000) proposes three types of assessment practice based on an investigation of primary school teachers' practices of assessment in the New Zealand context. These are:

> *Unit assessors*, who usually do not integrate assessment into their daily teaching practices. Assessment is used at the end of each school term or teaching unit to judge how much students have learned. Therefore, this type of assessment is well-planned and implemented mostly in more formal ways;

> *Head note assessors*, who believe that assessment is built into the teaching process. It is seen as a powerful tool to motivate students to keep learning. This type of teacher appears to use few or no documents to record students' performance but keeps much of the assessment information in his or her head; and

Integrated systematic assessors, like Gipps et al.'s (1995) systematic planners, tend to record assessment information systematically. Integrated systematic assessors use the information about students' learning in their teaching practice for improvement as a primary function of the information.

M. F. Hill (2000) provides a clear picture of these three approaches to assessment in the New Zealand primary school context, showing the complexity of primary teachers' conceptions of assessment.

Also in New Zealand, Brown (2003) proposes a model of teachers' conceptions of assessment to identify primary and secondary school teachers' beliefs about using assessment. By adopting a structural equation measurement model to analyse responses from a questionnaire, and using follow-up interviews, assessment conceptions of both primary and secondary teachers in the New Zealand context were examined. Participants taught subjects including English, mathematics, and science. Results indicate that teachers' conceptions of assessment include four sub-concepts: *improvement*, *accountability of schools and teachers*, *accountability of students* and *irrelevance*. Teachers' conceptions for improvement refers to teachers identifying student inadequacies in learning and remedying them in time. However, Brown (2003) claims that "there is little evidence that teachers actually consider their teaching as a possible object of change when students do poorly on assessment" (p. 29). Many teachers are found lacking knowledge and strategies to conduct effective assessment activities in classrooms.

Regarding teachers' conceptions of accountability, Brown describes two kinds of teachers' accountability conceptions: *school or society accountability*, and *student accountability*. Teachers with conceptions of assessment as school or society accountability tend to use assessment results to represent their work to the public as well as providing evidence to illustrate quality education and improvement resulting from schools and teachers. High-stakes tests with uniform standards are undertaken by teachers for this purpose. However, as Brown argues, there is a risk of sacrificing education quality by raising standards for increased accountability of teachers and schools. Regarding the conception of student accountability, Brown argues that, in the New Zealand secondary context, students are under pressure from NCEA examinations and therefore some teachers conceive students themselves as responsible for their performance. Finally, the *irrelevance* conception signals that some teachers believe that assessment is equivalent to tests, used for controlling students, rather than improving their learning and therefore classify it as irrelevant to the educational purpose.

Brown's (2003) study provides an understanding of different primary and secondary school teachers' conceptions of assessment in the New Zealand context at the time. He stresses that assessment practice will not be improved if teachers do not reveal and understand their own conceptions of assessment. Likewise, teachers' professional development will be less successful if their existing conceptions are overlooked.

Although there is a fairly extensive literature on teachers' assessment conceptions (Black & Wiliam, 1998, 2009, 2018; J. Chen & Brown, 2016; Cowie & Bell, 1999; Eyers, 2014; M. F. Hill, 2000; Nasr et al., 2020; Remesal, 2011), research focused on preservice teachers' assessment conceptions is less common. Due to the specific interest of this thesis in exploring changes in teachers' assessment conceptions from the preservice stage to the in-service stage, the next section reviews studies relating to preservice teachers' conceptions.

Preservice Teachers' Conceptions and Practices of Assessment

Studies of preservice teachers' assessment conceptions and practices are limited and appear to be influenced by the cultural context. Therefore, this section presents previous studies according to context, and focuses on the factors influencing conceptions and practices developed within these settings.

In the research area of preservice teachers' conceptions of assessment, Graham (2005) investigated the changes of preservice teachers' assessment conceptions and practices during a U.S. ITE programme. A total number of thirty-eight secondary English preservice teachers were asked to write down their responses to five assessment issues and provide related evidence of their assessment learning during the practicum. The results of this study show preservice teachers alter their prior beliefs about using assessment for grades, gaining confidence in adjusting assessment practices to address student's learning needs. Their changes in assessment conceptions and practices are highly influenced by two factors: preservice teachers' acquisition of assessment theories and the models that mentors provide during practicum. Graham (2005) argues preservice teachers need opportunities to deepen theoretical knowledge and work with experienced mentors who can bring theories into practice, in this way, their prior beliefs about assessment can be challenged in a meaningful way. Graham's (2005) work provides empirical evidence to show that preservice teachers refine their conceptualised assessment practices continually when given practice opportunities.

In contrast with Graham's (2005) findings, Karp and Woods (2008) found that preservice teachers' assessment practices are inconsistent with their understandings of assessment. In

their study investigating 17 U.S. preservice teachers' conceptions of assessment through surveys and interviews, preservice teachers reported having difficulties implementing effective assessment practices, given that factors such as specific school contexts restrict their assessment practices.

Also in North America, Coombs et al. (2020) propose three types of preservice teacher assessors: *eager*, *contemporary*, and *hesitant*. Using a quantitative approach, they examined U.S. preservice teachers' approaches to classroom assessment, self-perception of assessment competence, motivation for completing teacher education, and assessment education experiences. Eager assessors are preservice teachers who adopt standard assessment criteria to assess individual students' learning needs. This type of preservice teacher may not be able to adjust assessment approaches based on specific classroom contexts.

Some preservice teachers are described as hesitant assessors, as they conceive assessment as an isolated activity rather than a strategy within their pedagogy. Thus, they are unlikely to embed assessment into their teaching practice. Coombs et al. point out that this type of preservice teacher could benefit from teacher education with a structured approach to unpack their prior beliefs about assessment.

Contemporary assessors are found to be the largest percentage of participants, endorsing the belief that assessment is positively related to teaching and indicating they could gain enough capabilities of assessment through teacher education. Preservice teachers who hold such beliefs are more likely to engage with contemporary ideas regarding assessment such as formative assessment in their classroom practices, showing the success of teacher education in changing their conceptions and practices. However, as Coombs et al. point out, preservice teachers in primary schools have more motivation to try contemporary approaches of assessment than their secondary counterparts, given that preservice primary teachers may receive more school-level support to employ contemporary approaches in classroom teaching. In addition, preservice primary teachers are likely less stressed than their secondary counterparts by the pressure from high-stakes examinations. Coombs et al.'s study provides valuable knowledge about preservice teachers' conceptions of assessment. However, as they indicate, this study was conducted in the U.S. context, so more research is needed in other contexts.

In the Chinese context, Jiang (2015) conducted an ethnographic study to investigate how preservice teachers incorporate assessment knowledge into teaching practice and

communicate with other stakeholders during their 12-week practicum. By employing the approaches of interviews and observations, eight preservice teachers' assessment experience during the practicum was recorded. As Jiang (2015) found, preservice teachers' classroom assessment practices consist of two types: *the technical act of assessment*, which indicates preservice teachers' measurement of teaching and learning quality; and *the evaluative act of assessment,* in which preservice teachers assess student learning tacitly to improve teaching qualities. In the test-dominated assessment environment, both summative and formative assessments are interwoven in preservice teachers' classroom practices, and frequent summative assessment results are used to support student learning formatively. Preservice teachers keep adapting and refining what they have learned from teacher education into specific classroom contexts through methods of formative feedback upon test results to students. Jiang (2015) indicates that the process of adaptation not only involves the expansion of preservice teachers' existing knowledge but also includes adjustment of social, cultural, and affective factors.

Jiang's (2015) work sheds light on how preservice teachers adapt and deepen their conceptions and practices in the Chinese context. The results of this study show contextual factors from policy and school levels mediate preservice teachers' assessment development within teacher education programmes.

In the New Zealand context, there is no study about preservice secondary Chinese-heritage language teachers' conceptions and practices of assessment. In primary teacher education, however, there are some studies. Eyers (2014) investigated eight preservice primary teachers' assessment learning in teacher education in the New Zealand context. By employing a case-study approach, Eyers (2014) found that preservice teachers change conceptions significantly from negative to positive through teacher education. Their development of assessment conceptions and practices are influenced by prior beliefs, dispositions, coursework, and practicum experiences in teacher education. Similarly, Smith et al. (2014) examined preservice teachers' conception changes in assessment during 3 years of teacher education and found that preservice teachers can be prompted to learn assessment with opportunities to self-reflect during curriculum courses and practicum.

However, E. Hargreaves (2007) points out that many preservice teachers are still underprepared at the exit from teacher education. Factors such as short-term teacher education programmes limit preservice teachers' understanding and skills of assessment. Thus, more

studies are required to investigate changes of preservice teachers' conceptions, preservice Chinese-heritage language teachers in particular, during teacher education.

Beginning Teachers' Conceptions and Practices of Assessment

Beginning (or novice) teachers are a particular subset of in-service teachers; they are usually defined as teachers in their first 3 years of teaching (Flores, 2006). Farrell (2012) reviews previous literature and identifies beginning English-language teachers as "those who are sometimes called newly qualified teachers, who have completed their language teacher education program (including teaching practice [TP]) and have commenced teaching English in an educational institution (usually within 3 years of completing their teacher education program)" (p. 437).

While studies have shown that preservice teachers' knowledge and skills may not be strong enough to support them to create and communicate assessment, although they may espouse understandings and values of formative assessment at the exit from teacher education (Coombs et al., 2018; M. F. Hill & Eyers, 2014; Timperley et al., 2007), studies about beginning language teachers' assessment conceptions and practices in their early teaching careers are rare (Booms et al., 2017; Edwards, 2017; H. Xu, 2017), and no studies regarding beginning Chinese-heritage language teachers' conceptions and practice of assessment could be found. This may be because the research area of assessment capability is relatively new (DeLuca, Willis, et al., 2019; Edwards, 2017). Therefore, empirical studies regarding assessment conceptions and practices of beginning teachers (Coombs et al., 2018; Edwards, 2017; Edwards & Edwards, 2017; Nolen et al., 2011), and beginning language teachers (H. Xu, 2017) are reviewed. Furthermore, a longitudinal study of secondary Chinese-heritage language teachers' professional development in New Zealand is reviewed to gain understandings about Chinese-heritage teachers' teaching and learning experiences in the New Zealand context.

Coombs et al. (2018) adopt the Approaches to Classroom Assessment Inventory to compare preservice, beginning and experienced teachers' approaches to assessment. They found that teachers at different career stages hold different priorities for assessment. Beginning teachers (up to 5 years teaching) were found to prioritise summative assessment, and thus they are more likely to adopt a standardised assessment approach. Coombs et al. assert that beginning teachers who finish preservice teacher education programmes with knowledge of formative assessment seem to shift their conceptions and priorities to summative assessment after

entering the teaching profession, although preservice teacher education and professional learning are important in changing preservice teachers' understandings of fairness in assessment.

In a study regarding beginning language teachers' assessment conceptions, H. Xu (2017) observed the development of four beginning EFL teachers' assessment capabilities in a 3-year longitudinal study in the Chinese context. By employing an approach of qualitative data analysis, beginning EFL teachers' development of assessment capabilities are described as occurring in three stages. The first stage refers to their first year of teaching, where novice teachers learnt practical assessment techniques and built up their own assessment toolboxes. However, their understanding of planned assessment activities was found to be isolated from improvement of teaching. In the second stage, lasting for 2 years, novice teachers became more experienced in using designed assessment tools to serve teaching purposes. The participants were observed to use assessment information as feedback for teaching improvement. The third stage began at the end of the third year of teaching, where novice teachers were found to be more aware of using formative assessment information to improve teaching immediately. H. Xu also depicts a route of beginning Chinese EFL teachers' assessment learning in their first 3 years of teaching. The beginning teachers' personal learning autonomy and abilities were found to be critical in learning and promoting formative assessment in classrooms. Furthermore, as H. Xu suggests, beginning teachers' frequent use of unplanned formative assessment within classrooms shows their improved assessment awareness and capabilities of linking assessment and teaching together. However, as factors that influence novice Chinese EFL teachers' assessment learning are not discussed, the factors that triggered their changes of conception and practice of assessment are mentioned less in his study.

With a research interest in summative assessment literacy (capabilities to make use of summative assessments), Edwards (2017) traced eight beginning science teachers' development of summative assessment literacy from their entrance into a 1-year teacher education programme to the first 6 months of teaching in New Zealand secondary schools. Edwards describes a practical instrument, the Summative Assessment Literacy Rubric to observe and evaluate beginning teachers' assessment development at different stages. Through analysis of participants' responses at different points of learning and teaching, Edwards found that beginning secondary teachers' development of summative assessment literacy is individual and distinct from others. Personal factors (such as a knowledge base of

assessment) and contextual factors (such as ongoing opportunities to learn about assessment in schools) mediate beginning teachers' assessment learning. Furthermore, exams may be less fair when teachers are granted more initiative in designing and implementing NCEA internal examinations and thus hold more power to control the process of examinations.

Edwards' study illuminates how beginning secondary teachers mediate their ideologies about assessment within specific school contexts and the qualification system. However, as this study investigates the first 6 months of beginning secondary teachers' practices of summative assessment in secondary schools, how personal factors and contextual factors support them to develop summative assessment capabilities in the long term is less known. Furthermore, New Zealand's cultural make-up is diverse; beginning teachers from different ethnic groups are likely bringing their cultural backgrounds to bear on their teaching, and this may also influence their conceptions of assessment. Nevertheless, cultural influences on beginning secondary teachers' development of summative assessment capabilities are not obvious in this study.

In another area of beginning teachers' experiences, Nolen et al. (2011) explore novice secondary teachers' assessment learning in their first 2 years of teaching in the US. Assessment practices of seven novice teachers of different subjects are investigated. The results show that beginning teachers change their assessment practices by negotiating assessment tools and artefacts with the social world. However, it is not always the case that beginning teachers are willing to negotiate their assessment conceptions and practices with the social world. Mediating factors that contribute to their willingness to learn about assessment derive from their *individual agency*: "individual values, beliefs, identities, and relationships" (Nolen et al., 2011, p. 110). Nolen et al. (2011) argue that beginning teachers' prior experiences of assessment continually play a critical role in filtering and constructing their assessment practices. Even beginning teachers who were well prepared in teacher education were found to have difficulties in adopting advocated assessment practices when these practices contradict their prior experiences in assessment. Nolen et al.'s study sheds light on the complex decisions behind individual beginning teachers' assessment practice changes in the US context, showing the importance of understanding beginning teachers' prior experiences. More studies regarding beginning teachers' assessment conceptions are needed in other contexts.

Edwards and Edwards (2017) agree with Nolen et al.'s (2011) work, stressing that information from professional development programmes is filtered and interpreted through novice teachers' existing preconceptions, some of which are impacted by their cultural background. By conducting a longitudinal case study, Edwards and Edwards (2017) explore the professional growth of a novice Māori secondary teacher from his 1-year teacher education through the first year of teaching in a New Zealand English-medium secondary school. The results show that the beginning teacher's adaption of assessment approaches into a specific school culture was challenging. The participant's choices of assessment tools and assessment practices were more likely influenced by his prior experiences as well as his cultural background. Moreover, as a new teacher, with a bicultural background, he confronted struggles when his conceptions contradicted advocated assessment practices. The findings suggest that beginning teachers need more time to improve their assessment literacies through constant interactions of their own conceptions with the social world. In addition, Edwards and Edwards argue that, in the New Zealand secondary sector, novice teachers face difficulties in refining their assessment practices when working in the tight schedules of preparing students for NCEA. Edwards and Edwards' work reveals that novice teachers are challenged to shift assessment conceptions and practices in their first years of teaching in the New Zealand context. However, their work only reports the journey of one novice teacher from the Māori ethic group building up teacher identity from preservice stage to in-service stage; there are limited studies tracing beginning teachers from other ethnic groups to learn about assessment conceptions and practice in the early years of teaching.

Although not in the assessment realm, Sun (2010) conducted a longitudinal study to record and compare three Chinese language teachers' professional development of personal practical knowledge in the New Zealand context. All the participant teachers were from New Zealand secondary schools, two of them were Chinese-heritage language teachers, and one of them was non-native Chinese teacher. Research methods included interviews and classroom observations. The results show that traditional Chinese culture significantly influenced the two Chinese-heritage teachers' teaching conceptions and practices. Firstly, both experienced and novice Chinese-heritage teachers faced challenges from the distinctive cultural differences between the East and the West. Chinese-heritage teachers brought with them conceptions and standards which were formed in Confucian-heritage cultures (CHCs), and these standards and conceptions still impacted their teaching and assessment practices. Specifically, the Chinese-heritage teachers sometimes implemented teacher-centred

approaches as well as evaluating student performance and achievement, which was inconsistent with advocated teaching and assessing practices in the New Zealand context. Secondly, Chinese-heritage language teachers were found to be lacking in confidence to implement effective assessment practices in the classroom, given that teachers' assessment practices in the Chinese context need to follow the textbooks closely as well as meet the requirements of high-stakes examinations. In contrast, in the New Zealand assessment context, teachers are freer than their counterparts in China to choose different types of assessment. Therefore, these novice teachers experienced difficulties in teaching and lacked assessment resources in their first year of working.

Sun's (2010) work highlights the importance of considering novice Chinese-heritage teachers' cultural background in teacher education and professional development, as he argues that the influence from CHCs on their professional development is stronger than their individual agency. Sun suggests that novice Chinese-heritage teachers need more opportunities in professional development programmes to reveal and develop their own practice. However, despite the fact that his study focuses on secondary Chinese-heritage teachers' professional growth, there is relatively little discussion in his study about changes to Chinese-heritage language teachers' conceptions and practices in New Zealand secondary schools.

With increasing research emphasis on teachers' assessment capabilities, investigations into the changes in assessment conceptions and practices of beginning secondary Chinese-heritage language teachers, during their early teaching careers, could enrich the existing literature regarding beginning teachers' assessment learning and their professional development of assessment capabilities.

Factors that Influence Conceptions and Practices of Assessment

Studies have shown that teachers are challenged to implement effective assessment, especially effective formative assessment, in specific school and classroom contexts (Cooper & Cowie, 2010; DeLuca, Willis, et al., 2019; M. F. Hill, 2011, 2016). Cooper and Cowie (2010) indicate that teacher assessment is not only a personal challenge but is also influenced by contextual factors. Carless (2005) clarifies these influences by proposing three levels of factors that influence teachers' changes of practice and conceptions of assessment: personal, micro- and macrolevels. Personal-level factors include teachers' personal understandings and principles, and consistency between existing beliefs and practice; microlevel factors indicate influences from school context, such as internal school support, views of parents; Macrolevel factors

include influences from a political level, such as a reform climate, learning and assessment, and the role of high-stakes examinations. Based on findings of previous research, the following subsections review factors that may contribute to changing beginning secondary Chinese-heritage language teachers' conceptions and practices of assessment using these three levels of factors. Specifically, studies regarding personal factors of beginning teachers' prior experiences of assessment are firstly presented. Then, microlevel factors including beginning teachers' assessment experience in teacher education and their assessment experience in school communities are discussed in the following sections. Finally, literature of New Zealand assessment contexts and CHCs is reviewed respectively as macrolevel factors which may have an impact on beginning Chinese-heritage language teachers' assessment.

Preservice and Beginning Teachers' Prior Experiences of Assessment

Teachers' prior and current experiences of assessment both result in changes of conceptions and practices (Black & Wiliam, 1998; Brookhart, 2017; Crossman, 2004; Ell et al., 2012; Heng & Hill, 2018; M. F. Hill, 2016; Timperley, 2007). Crossman (2004) refers to teachers' prior experience of assessment as their "past histories" (p. 585). In an investigation of preservice teachers' conceptions of assessment in the Australian context, Crossman (2004) adopts approaches of semi-structured interviews and observations to explore factors that influence preservice teachers' conceptions of assessment. Results show that some participants were reluctant to enrol in the assessment courses, given that they had understood assessment in a negative and surface way in the past. However, as Crossman asserts, there is no direct evidence to prove the connections between preservice teachers' prior and current conceptions of assessment besides participants' personal reflections.

In New Zealand, M. F. Hill, Gunn, et al. (2014) examined preservice teachers' conceptions of assessment at different points as they completed a 3-year ITE programme. By employing the approaches of questionnaires, individual interviews and focus group interviews, 720 primary and 120 early childhood preservice teachers' assessment learning during the 3-year ITE programme were recorded. The results show that the preservice teachers held broad understandings about different assessment purposes and the relationship between these purposes at the entry of the ITE programme. During their learning about assessment in the ITE programmes, this cohort of preservice teachers shifted their understandings of assessment from a student perspective to a teacher perspective. M. F. Hill, Gunn, et al. (2014) find that preservice teachers deepened understandings of assessment and felt more confident at using assessment as they began to teach in schools. Nevertheless, they had limited understanding to

engage students in classroom assessment throughout their learning during the ITE programme. M. F. Hill, Gunn, et al suggest that school leaders and mentors should continue to support beginning teachers' assessment learning as they begin to teach. Furthermore, M. F. Hill, Gunn, et al. (2014) suggest that preservice and beginning teachers need to take initiative in their own assessment practices and assessment learning. In this way, they can become more assessment capable to support children as they begin teaching. M. F. Hill, Gunn, et al.'s (2014) study confirms that preservice teachers hold a certain level of assessment understanding at the entry of teacher education programmes, and their personal motivation about learning assessment influences their continued development of assessment capabilities. However, their study focuses on preservice primary and early childhood teachers' assessment conceptions during a 3-year ITE programme; how secondary language teachers learn about assessment during ITE programmes is not addressed.

Timperley et al. (2007) reviewed previous studies on teachers' learning processes through a best evidence synthesis, based on which they summarise three major processes of teacher learning: 1) cueing and retrieving prior knowledge; 2) becoming aware of new information/skills and integrating them into current values and belief systems; 3) creating dissonance with current position (values and beliefs). As Timperley et al. (2007) suggest, teachers' practices are strongly influenced by their prior experience. The more dissonance between their prior experience and current positions, the more challenges exist for teachers to adjust their assessment practices.

Similarly, Y. Xu and Brown (2016) synthesise empirical studies regarding teachers' conceptions and practices of assessment and provide an understanding of teachers' assessment capability in practice. As they stress, it is essential to investigate teachers' existing knowledge and beliefs to find out what factors promote or hinder teachers' professional learning, as new knowledge of assessment is always filtered and interpreted by teachers' conceptions of assessment. Therefore, the outcomes of professional assessment training are always "offset" (p. 155) by teachers' existing knowledge and beliefs of assessment.

The Influences from Assessment Education Programmes

There is a growing body of research regarding teacher conceptions of assessment and their relationship with assessment education. However, results are inconclusive. Some research has found that preservice teachers' conceptions of assessment can change dramatically with sustained attention through teacher education (M. F. Hill et al., 2017; Levy-Vered & Nasser-

Abu Alhija, 2015; Smith et al., 2014), while other studies have found preservice teachers may not change their conceptions during learning in teacher education programmes (Brown, 2008; DeLuca & Volante, 2016; Hammerness et al., 2005; Karp & Woods, 2008; Wubbels, 1992). On the one hand, preservice teachers have been found to hold mostly negative views about assessment and to conceive of assessment in mostly summative ways on entry to teacher education programmes (M. F. Hill et al., 2012). Following teacher education with a focus on assessment, preservice teachers' attitudes move to become more positive towards the use of formative assessment. Levy-Vered and Nasser-Abu Alhija (2015) conducted a quantitative study in the Israeli context to test the relationship between assessment literacy, assessment conception, and self-efficacy. A total of 327 Israeli beginning and in-service teachers responded to three questionnaires and a test in assessment literacy. The results show that beginning teachers who have acquired knowledge and skills from teacher education are more assessment literate than their in-service counterparts.

In the Canadian context, DeLuca and Klinger (2010) employed questionnaires to examine 288 both primary and secondary preservice teachers' assessment confidence levels, understandings and philosophies in a teacher education programme. They compared responses from the two cohorts of student teachers who had and had not attended the assessment course. Their findings support the idea that assessment courses correlate to student teachers' assessment confidence in terms of theories and practices. Also, constructing assessment courses with direct instructions on assessment skills such as reports on student achievement could benefit student teachers' assessment learning. They also found that preservice teachers built up their assessment confidence through practicum practices when an assessment course was not compulsory for them, and thus preservice teachers' development of assessment was dependent on associate teachers' capabilities of teaching and assessment.

Similarly, in the New Zealand context, M. F. Hill, Ell, et al. (2014) explored the influence of an assessment course on a cohort of preservice primary teachers' development of assessment capabilities. By analysing assessment course booklets, and teacher educators and preservice teachers' conceptions of assessment at different points in their learning in a 3-year teacher education programme, M. F. Hill, Ell, et al. found that this cohort of preservice teachers demonstrated the assessment competencies required by the graduating teacher standards[2]

---

[2] M. F. Hill, Ell, et al. (2014) listed four standards of assessment competence that graduating teachers need to achieve: (1) use evidence to promote learning; (2) systematically and critically engage with evidence to reflect and refine their practice; (3) gather, analyse and use assessment information to improve learning and inform

(New Zealand Teachers Council, 2007), and showed confidence in using assessment to enable student learning as they began teaching.

On the other hand, a number of studies have found that short-term training on assessment has a limited effect on changing teachers' existing conceptions, and professional development may not resolve teachers' misuse of assessment (Brown, 2008; DeLuca & Volante, 2016; Hammerness et al., 2005; Karp & Woods, 2008; Wahl et al, 1984; Wubbels, 1992). Karp and Woods (2008) examined 17 preservice teachers' conceptions of assessment at different stages in their physical education programme in the U.S. context. The findings suggest that while preservice teachers started to absorb knowledge from teacher education programmes, there was not enough time for them to revise their previous conceptions and improve their teaching practices. Brown (2004) explains that, due to the complexity of assessment conceptions, it is essential to gain an understanding of teachers' prior experience of assessment, which may influence the effectiveness of teacher education.

Also, the effectiveness of teacher education programmes is closely related to support from associate teachers or mentors to practise assessment learned through the coursework (Edwards, 2017; Eyers, 2014; Darling-Hammond, 2006; Graham, 2005; McDonald, 2004). Associate teachers' beliefs about teaching, assessment and learning impact their supervision of preservice teachers' classroom practice. As Swell et al. (2017) suggest, "being an effective classroom teacher does not automatically imply being a skilled AT [associate teacher]" (p. 23). McDonald (2004) invited four associate teachers and six preservice teachers in the New Zealand context to discuss factors that contribute to successful supervision by associate teachers. The findings suggest seven qualities that associate teachers should be equipped with to guide preservice teachers' practice: personal pedagogy, role models, reflection, feedback, communication, personal professional qualities, and professionalism. As McDonald suggests, encouragement and effective feedback can effectively assist preservice teachers to improve their teaching, and ultimately benefit student learning. In contrast, large workloads for preservice teachers and little feedback and guidance from associate teachers can lower preservice teachers' confidence in using assessment. Also, associate teachers may provide ineffective supervision when they receive insufficient professional development.

---

planning; (4) know how to communicate assessment information appropriately to learners, their parents/caregivers and staff.

Also in the New Zealand context, Eyers (2014) examined eight preservice primary teachers' assessment learning during a 3-year teacher education programme. While earlier research emphasised that associate teachers' formative and summative assessment practice could have a great influence on preservice teachers' assessment conceptions, Eyers's study indicates that associate teachers' limited assessment capabilities may hinder preservice teachers' attempts to put their assessment knowledge into classroom practice.

Continuous Learning of Assessment in Professional Learning Communities

Teacher education may be neither adequate nor enough to support beginning teachers to adapt their knowledge and skills into "the real world" (DeLuca et al., 2012; DeLuca et al., 2018; DeLuca, Willis, et al., 2019; Farrell, 2006; Northfield & Gunstone, 1997). As Farrell (2006) explains, "learning to teach is often viewed as being a highly situated, highly interpretative, and idiosyncratic activity" (p. 211); teachers' assessment learning is not only dependent on their own beliefs, but is also influenced by the microlevel factors, such as school assessment contexts or internal and external school support. Beginning teachers' experiences in the first years of teaching and assessing are usually involved with reconstruction of practices within a specific school culture (Bullough, 1997; Farrell, 2006, 2012). However, research has found that schools are yet to build up formative assessment communities to support beginning and even experienced teachers to learn assessment continuously (Brown, 2011; Chandan, 2018; DeLuca et al., 2012; M. F. Hill, 2016; Nolen et al., 2011). M. F. Hill (2011) investigated factors enabling secondary school teachers to develop effective formative assessment use in the New Zealand context. By conducting focus group interviews with principals and lead teachers (in charge of assessment) in three secondary schools that had established strong formative assessment practices within their school community, she found that enabling formative assessment across schools and building up an assessment-for-learning community required effort from principals, senior staff and teachers to integrate professional learning about assessment within specific school cultures. In particular, M. F. Hill argues that building up assessment learning communities is found to be challenging in New Zealand secondary schools. This is because secondary teachers are likely to prioritise summative conceptions of assessment and student accountability due to the pressure from high-stakes qualification assessment systems and examinations as well as the micropolitics in subject departments.

In contrast, Brown (2011) compares New Zealand primary and secondary in-service teachers' conceptions of assessment through their responses to the teachers' conceptions of assessment and finds that teachers in both primary and secondary sectors hold similar conceptions of

improvement are in the New Zealand context. However, from their examination of secondary teachers' conceptions of assessment, Yates and Johnston (2018) argue that student's conception of assessment accountability in the New Zealand context is positively correlates to assessment for improvement. Secondary teachers in the New Zealand context may, like other teachers who work in high-stakes examination countries, put a strong emphasis on student achievement in examinations/assessment for qualifications.

Also, in the New Zealand context, Cooper and Cowie (2010) stress the importance of external support for secondary teachers' continuous learning about formative assessment. They conducted an interventional study, tracing three secondary teachers' changes of assessment conceptions and practices with support from two university researchers. The results show that the secondary teachers developed shared understandings and principles of assessment and thus strengthened their formative assessment practices. Cooper and Cowie (2010) suggest that formative assessment innovations could be successfully implemented in schools when personal factors and microlevel factors, such as internal school and external support, communicate and interconnect. In this way, teachers could become part of formative assessment innovations.

Second-Language Teaching and Assessment in the New Zealand Context
As Carless (2005) points out, language teachers' assessment practice is not only influenced by personal and school-level factors, but also by macrolevel factors such as the assessment reform climate and the societal teaching, learning and assessment culture. The impact of relevant government or quasi-governmental agencies promotes or hinders the extent to which formative assessment is implemented in classrooms.

In the New Zealand context, teaching and learning Chinese language is an optional subject in schools, and seen as a second (L2) or other language for students to learn. *learning languages* (Ministry of Education, 2007), was introduced as a learning area in the revised *New Zealand Curriculum* in 2007, demonstrating the intention to strengthen the importance of learning languages in schools (East, 2008; Ministry of Education, 2007; Ward & East, 2016). Students from Years 7–10 (11 to 14 years of age) are encouraged to sample one or more additional languages or so-called "taster" courses for a short period of time. Through learning in such short-term courses, students can decide whether they would like to continue learning additional languages. However, the development of L2 language programmes in secondary schools did not go smoothly in the New Zealand context (Ashton, 2018; East, 2008, 2016,

2018; Ward & East, 2016). Ward and East (2016) employed a mixed-methods study to investigate stakeholders' perspectives of language-teaching programmes. Seven curriculum leaders and principals within four New Zealand secondary schools were interviewed, and 117 in-service teachers completed an online survey. The findings suggest four factors influence the development of L2 language programmes in secondary schools. First, both principals and leaders agreed that secondary schools lack qualified L2 language teachers. In New Zealand, L2 language learning is optional for students. L2 teachers' teaching quality strongly influences students' enrolments and thus the retention of such programmes varies from school to school. Furthermore, L2 teachers reflected that they have to compete with STEM (science, technology, engineering and mathematics) disciplines which seem to be more popular in the crowded curriculum. Third, responses from language leaders highlighted that their work did not receive sufficient support from their colleagues and the community. However, principals reflected that the community wants students to learn languages as an additional skill for future careers. Ward and East (2016) indicate that the status of the less valued second languages in New Zealand society impacts students' motivation to learn languages. Finally, L2 teaching is challenged in secondary schools given that L2 learning is not compulsory. Ward and East (2016) suggest that making learning languages compulsory could enhance the perceived value of learning languages in secondary schools and support the development of L2 programmes financially.

Ward and East's (2016) study is valuable for understanding the status of L2 programmes in secondary schools and the factors that may hamper the success of such programmes in the New Zealand context, even though the status of the development of Chinese language as a subject was not the focus of the study.

Jones (2014) explores the development of L2 programmes in New Zealand secondary senior levels. Through the analysis of statistics about nationwide enrolments in L2 programmes in Year 13, as well as an analysis of curriculum documents regarding learning languages, Jones points out that, between 2009 and 2010, in contrast with the policy encouraging learning languages at senior levels and high participation rates, fewer students enrolled in senior L2 language courses in 2010. As a result, less than 10% of secondary school students chose to study a second language, and only 20% of New Zealand secondary schools have L2 language classes for Year 13 students. Jones argues that learning second languages is a benefit for multilingual cultures in New Zealand, while the development of L2 programmes needs more support at policy level, such as setting goals for secondary school leavers to be proficient in

L2. Furthermore, Jones suggests that authorities should ensure the NCEA credits in L2 assessments and in other subject assessments are equally valid. Finally, Jones suggests secondary school leaders should make positive changes to the timetabling of L2 so that learning L2 more accessible for students. While Jones's study is valuable to understanding the situation of the development of L2 programmes in secondary schools, less is mentioned about L2 teachers' teaching and assessment in New Zealand secondary schools.

Ashton (2018) explores secondary language teachers' views of teaching multilevel classes through their responses in interviews and a survey. She found that L2 teachers face challenges from the undervaluing of language learning at secondary school level, in the broader community and at government policy level. The findings show that language subjects are marginalised by curriculum decisions in schools. In addition, over half of the participating teachers reported that they received very little support from their school as well as limited opportunities for professional development. Ashton's study reveals the issue that L2 subjects are less valued in New Zealand secondary schools, although her focus is not on the subject of Chinese language.

East (2008, 2012) further discusses New Zealand language policies and sociocultural influences on the development of L2 language programmes in schools. By comparing educational policies in the New Zealand and English contexts, East (2008) points out that learning L2 languages is only offered to students in New Zealand from Year 7–10 as a suggestion, whereas learning languages in English schools is a compulsory core subject. The potential disadvantage is that L2 language programmes in New Zealand may receive unequal development opportunities in comparison with compulsory subjects (such as science or social studies). At the same time, problems around both financing and staffing of L2 language programmes have been identified as "potential barriers to success" (p. 120).

Furthermore, East (2008) points out that the low uptake of L2 language learning in New Zealand secondary schools is related to the fact that learning languages is less valued in New Zealand. He argues that teaching and learning L2 requires a cultural change, in which young people need to shift their conception of English as the only essential language in New Zealand. East's (2008) study places New Zealand L2 education in the international context and demonstrates the importance of supporting L2 language teaching and learning at the policy level and sociocultural level. In the report for the New Zealand and China Council, East (2018) points out that the development of L2 language programmes in New Zealand schools also

faces the challenge of a lack of qualified teaching staff, a crowded curriculum, timetabling, and a lack of support from government policy in New Zealand.

The Influence of Chinese-Heritage Cultures on Chinese-Heritage Teachers' Conceptions and Practices of Assessment

Language teachers' conceptions and practice of assessment are inevitably marked by their cultural background (Brown & Remesal, 2012). Borg (2015) stresses the importance of researchers looking at the influence of sociocultural factors on teachers' conceptions and practices. For Chinese language teachers who have had prior experience teaching and/or learning in both the New Zealand and Chinese education contexts, conceptions and practices of assessment could be more complicated, given that their conceptions and practices of assessment are likely to be influenced by both New Zealand and Chinese cultures. Carless (2012) described CHCs as an educational culture that includes exam-oriented, teacher-centred transmission-style pedagogy; and high parental investment. Song (2016) indicates that CHCs are rooted in Eastern countries such as China, Japan, Korea, and Singapore. China is historically examination-oriented and has been for more than 1,000 years. A high-stakes examination-dominated environment not only strengthens the selective purpose of assessment but also deeply constrains stakeholders' use of assessment to improve learning. Carless (2012) proposes a programme of formative use of summative testing for second-language teachers in a Hong Kong secondary school which aims to improve student learning by providing training for teachers to improve their assessment literacy. Results show that teachers' assessment practice is still constrained by the accountability requirements of the education system, especially under the pressure of highly competitive university entrance exams.

Kennedy (2016) concludes four characteristics of teachers' conceptions of assessment in CHCs: First, teachers in CHCs hold different conceptions of assessment from their counterparts in Western societies. They are likely to use assessment to enhance teaching and learning as well as social connections. Furthermore, students' examination results are closely related to teacher and school accountabilities in CHCs. In addition, as Kennedy (2016) points out, teachers' conceptions in different contexts (such as in Hong Kong and in mainland China) are different, showing social and cultural contexts interact in CHCs and prompt teachers to conceive assessment in different ways.

To understand the influences of CHCs on learners and teachers' conceptions and practices, Perez and Shin (2016) investigated 80 students' responses to teaching and learning styles in

CHCs by employing methods of interviews and questionnaires. They found that teachers' roles in CHCs are centred around four *H*s: hierarchy, harmony, hard work, and humanism (see Table 2.1).

Table 2.1

Practical Applications of Chinese-Heritage Cultures in Teaching

Practical applications	
Hierarchy	The hierarchy of relationships throughout the learning process is essential
Harmony	In an effort to preserve group harmony, CHC students display learning styles conducive to collaborative learning
Hard Work	The Confucian work ethic and sense of sacrifice is so well-known that it has become a stereotype
Humanism	Instructors can still help bring about self-actualisation by fostering mutual respect in the classroom

Note. Adapted from Perez and Shin (2016, p. 4)

Perez and Shin explain that teachers working in CHCs are expected to deliver knowledge and interact less with students, as teachers are seen as the model of knowledge (Hierarchy). Furthermore, teachers are also responsible for keeping group harmony in school communities and thus they are "allowing group concerns to supersede their own" (p. 4).

Extensive studies further discuss the influences from the Chinese examination culture on both preservice and in-service Chinese teachers' assessment conceptions and practices (J. Chen & Brown, 2016; Davey et al., 2007; Gu, 2014; Liu et al., 2016). Davey et al. (2007) reviewed the influences of Gaokao—the high-stakes university entrance examination system—in the Chinese context. Through the analysis of previous literature and statistics regarding rates of student enrolment in one Chinese university in 2004, they claim that Chinese students and teachers are under pressure to excel in Gaokao due to the fierce competition for entry to universities. Therefore, teachers start preparing students for Gaokao at an early age; both students and teachers' assessment experience is shaped and centred around the purpose of achieving better scores in Gaokao.

By employing approaches of interviews and observations, Gu (2014) observed an experienced Chinese teacher teaching and using assessment in a top school in Beijing. He found that despite the published curriculum (Ministry of Education in China, 2010) that encourages cultivating students' "knowledge, skills and well-roundedness (quality)" (p. 287),

implementing the curriculum in the secondary school sector is still challenging (Gu, 2014; Zheng, 2013, 2015). This is because student achievement in high-stakes examinations is still seen as an important indicator of the accountability of schools and teachers. Thus, Gaokao is still dominant in deciding teaching and learning strategies. Formative assessment activities are only used as supplementary to teaching content or activities to prepare students for exams (Gu, 2014).

Carless and Lam (2014) explore practical strategies of adapting formative assessment practices in the CHC context. In a qualitative exploratory study in the Hong Kong context, they observed two primary English teachers' formative assessment practice in preparing students for an upcoming test. The participant teachers adopted two formative assessment strategies: students' creation of their own mock test papers and peer feedback in a test follow-up. They found that teachers' assistance with students' preparation of the mock test papers could encourage students to develop ownership of the test-preparation process. With regard to peer feedback on test papers, Carless and Lam found that such a formative strategy seems more promising in the CHC context, given that the "collectivist orientation of student learning may be as strong as, or stronger than, the more individualistic perspective of examinations as competition" (p. 175). Lam (2016) claims that formative assessment practices is inevitably adapted in the CHC context where the instrumental view of learning is dominant; Carless and Lam (2014) suggest that adapting formative assessment strategies in test-preparation and for test follow-up processes is promising in the CHC context. However, they point out that integrating formative and summative assessment could be dangerous, given that students may put their emphasis on achieving better scores in tests rather than improving learning.

Liu et al. (2016) argue that the influence of high-stakes examinations on teachers and students' assessment conceptions is not unique to the Chinese context. In a quantitative comparative study of preservice Chinese and U.S. teachers' conceptions of classroom assessment ethics, 173 preservice teachers in the US and 174 preservice teachers in China responded to a designed survey regarding classroom assessment and assessment ethics. Liu et al. found that preservice teachers from both the US and China agreed that teachers prioritise helping students achieve better scores in exams. The responses from participants show high agreement that teachers need to prepare students for high-stakes tests and exams. Through the preparation of tests, teachers can find out the non-achievement factors that contribute to students' performances in tests, and thus gain more information about students' mastery of teaching content. Furthermore, Liu et al. (2016) assert that preservice Chinese teachers are likely to

encounter difficulties in learning and teaching in U.S. classrooms, as they found that preservice teachers in both the US and China show high alignment with their own assessment culture. Therefore, they suggest teacher education programmes provide more opportunities for preservice Chinese teachers to raise awareness of cultural differences between different education and assessment contexts. Liu et al.'s (2016) study is important to understanding the difficulties that Chinese-heritage teachers may encounter when they learn and teach in a different educational context; nevertheless, the challenges for Chinese-heritage teachers' assessment learning in other education contexts were not well addressed.

From a different perspective, behaviourist conceptions of learning may also usefully explain CHCs teachers' conceptions and practices. Carless (2012) explains that behaviourist conceptions of learning influence teachers "to denote a school of strategies that involve breaking learning into smaller steps, pattern practice and correction or reinforcement" (p. 12). Compared with the constructivist and sociocultural conceptions of learning underlying the formative approaches to teaching widely promoted in the New Zealand educational context for decades (Bell & Cowie, 2001; Bell & Gilbert, 1996; Garbett, 2011; McPhail, 2016), teaching, learning and assessment practices in CHCs seem to be still influenced by behaviourist conceptions (Corcoran, 2014; Zheng, 2013). Carless (2012) asserts that mechanical practising stemming from behaviourist beliefs about learning is still popular in language-teaching processes. Teachers who adopt such didactic methods are more likely to frequently assess student-learning progression by tests (James, 2006; Shepard, 2000), their feedback is usually corrective (Tunstall & Gipps, 1996), and students are required to engage in further instructions and practices (James, 2006).

In contrast, constructivist theories situate learning as an active engagement in a social context (Shepard, 2000). A focus of this view is that any new knowledge is constructed from the prior understandings (Bransford et al., 2000; Gardner, 2006; Shepard, 2000). The difference between experts and beginning teachers is that experts are more competent in using metacognition skills, such as self-monitoring and self-regulation, to organise new knowledge and structures. Guided by this approach, teachers use classroom assessment, formative assessment in particular, to help students to improve learning achievements as well as learning processes (Garbett, 2012; Gardner, 2006; Shepard, 2000). Nevertheless, Lam (2016) asserts that formative assessment practice responds to a constructivist approach of teaching and learning; therefore, direct adoption of formative assessment practice is likely to encounter difficulties due to product-based instruction being dominant in the CHC context. Lam (2016)

points out that CHC language teachers are likely to experience frustration at reconciling the discrepancy between ideal teaching and assessment practice and the reality of accountability demands. He suggests that CHC language teachers need more time and opportunities to reflect their existing beliefs and experience of using assessment to support student learning in the CHC context, where tests still dominate in teaching and learning.

In summary, this section presents factors that may influence beginning Chinese-heritage language teachers' assessment conceptions and practices. While research has demonstrated that beginning teachers' personal beliefs impact their assessment learning, microlevel factors such as school assessment cultures, internal and external school support; and macrolevel factors such as the development of L2 programmes in New Zealand, as well as beginning Chinese-heritage language teachers' Chinese-heritage cultural background, have been discussed in this section. Next, previous research regarding frameworks of assessment practices is reviewed to describe a conceptual framework of this study.

Theoretical Frameworks of Assessment Processes

In this section, frameworks regarding teacher assessment processes are examined. Previous studies have provided frameworks to help understand teachers' assessment practices in depth. For example, frameworks have been developed depicting planned and unplanned assessment activities (Cowie & Bell, 1999; C. Turner & Purpura, 2015; Purpura & Turner 2018), formative and summative assessment (Black & Wiliam, 2018; Harlen, 2007), and the process of assessment (Dickins, 2001; Harlen, 2007; K. Hill & McNamara, 2012; Purpura & Turner, 2018). Below, previous research into teachers' assessment practices is reviewed in two aspects: processes (planning, eliciting, interpreting and using assessment), and purposes (formative and summative assessment). Then, a conceptual framework of teacher assessment for this study is described.

Frameworks of the Process of Language Teacher's Assessment Practice

Three frameworks were found in recent literature to describe the process of language teachers' planned assessment activities. Dickins (2001) explores and proposes four stages that language teachers move through to make assessment-related decisions; K. Hill and McNamara (2014) provide both language teachers' assessment processes and conceptions behind their assessment decisions. In addition, Purpura and Turner (2018) propose the process of planned assessment in language classrooms through three different phases: before, within and beyond assessment protocols, which is helpful to understand the chain of reasoning in classroom.

Dickins' Framework. Deriving from a larger study investigating primary English-as-additional-language teachers' assessment practices in England and Wales, Dickins (2001) analyses the data of assessment practices from one classroom teacher and two support teachers and identifies assessment processes as four decision-making stages (see Figure 2.4).

Figure 2.4

Dickins's Framework of Processes and Strategies in Classroom Assessment

Stage 1: Planning

- Identifying the purpose for the assessment? (why?)
- Choosing the assessment activity (how)
- Preparing the learners for the assessment
- Who chooses/decides for each of the above?

Stage 2: Implementation

- Introducing the assessment (why, what, how)
- Scaffolding during assessment activity
- Learner self-& peer monitoring
- Feedback to learners (immediate)

Stage 3: Monitoring

- Recording evidence of achievement
- Interpreting evidence obtained from an assessment
- Revising teaching and learning plans
- Sharing findings with other teachers
- Feedback to learners (delayed)

Stage 4: Recording & Dissemination

- Recording & reporting progress towards NC
- Formal review for LEA or internal school purposes
- Strategies for dissemination of formal review of learners

Note. Adapted from Dickins (2001, p. 435)

According to Dickins, the four-stage assessment process focuses on describing daily classroom assessment activities in which teachers take the initiative to make assessment decisions, and thus the three teachers' assessment strategies in each stage are listed. Dickins

stresses that effective assessment does not necessarily include all four stages; teachers need to make decisions based on their assessment purposes.

Dickins' framework provides an overview of the stages a teacher might go through using classroom assessment, and the assessment strategies they might use to recognise teachers' assessment practices. However, this framework merges both formative and summative assessment purposes within the four stages, and thus it remains unclear how the process meets each purpose.

K. Hill and McNamara's Framework. Drawing upon three essential stages of classroom assessment (*elicitation*, *interpretation*, and *use*), K. Hill and McNamara (2012) developed a framework to describe language assessment in the classroom, meeting both formative and summative purposes (see Figure 2.5).

Figure 2.5

K. Hill and McNamara's Framework of Classroom Assessment

1. What do language teachers do?	
1.1. Planning Assessment	Is there planning for assessment? How detailed is planning? What is its intended relationship to instruction? How does it relate to external standards and frameworks?
1.2. Framing Assessment	Is assessment made explicit to learners? How is this done?
1.3. Conducting Assessment	What opportunities does the classroom provide for assessment? Does assessment tend to focus on the class, group/pairs of students or individuals?
1.4. Using Assessment Data	How is assessment-related information used?
	Teaching
	Learning (feedback) — Person-referenced
	Task-referenced
	• Confirmatory
	• Explanatory
	• Corrective
	Reporting
	Management
	Socialization
2. What do teachers look for?	
What information about valued enterprises, qualities and standards is available?	
2.1. In Advance	in written/verbal instructions and/or assessment rubrics?
2.2. In Feedback	in written and/or verbal feedback?
2.3. In Reporting	in reporting deliberations and/or in written reports?
3. What theory or 'standards' do they use?	
3.1 Teacher Theories & Beliefs	What does the data reveal about teachers' beliefs about • the subject or content area • second language learning and teaching, and • the nature of assessment?
4. Do learners share the same understandings?	
4.1. Learner Theories & Beliefs	What does the data reveal about learners' beliefs about • second language learning, and • the nature of assessment?

Note. Adapted from K. Hill and McNamara (2012, p. 425)

This framework raises three crucial questions (What do teachers look for? What theory or "standards" do they use? Do learners share the same understandings?) for researchers when they make sense of language teachers' classroom assessment practices. K. Hill and McNamara investigated primary and secondary Indonesian language teachers' assessment conceptions in the Australian context. The results confirmed the processes of assessment (planning, framing, conducting and using assessment) in their proposed framework. K. Hill and McNamara argue that while previous literature discusses teachers' strategies or stages of assessment, there is no comprehensive framework to inquire into processes of teachers' assessment practices. Their work provides a comprehensive model to investigate and understand language teachers' classroom assessment practices. Nevertheless, their framework discusses teachers' planned assessment activities in classrooms, and the process of teachers' unplanned assessment activities remains unclear.

Purpura and Turner's Model. Purpura and Turner (2018) propose three models based on Carless' framework of LOA. The first describes the process of teacher assessment. In their framework, a completed classroom assessment event should include three phases: before the assessment, within the assessment protocol, and beyond the assessment protocol but within the event. Within each phase, there is a chain of reasoning as below (see Figure 2.6):

Figure 2.6

Purpura and Turner's model of The Assessment Chain of Reasoning

The Assessment Chain of Reasoning	
Before the assessment	(explicit/internalized) Design considerations → Development
Within the assessment protocol	Elicitation (instrumentation or naturalistic)→ Data (evidence) → Inferences (interpretations) → Decisions→
Beyond the assessment protocol but within the event	Actions (e.g., feedback) → Follow-through (elicitation time 2) → Efficacy (e.g., movement or partial movement in L2/topical system; learning;) → Beneficence (+ conseq.)

Note. Adapted from Purpura and Turner (2018, LTRC conference)

In their framework for assessment, the components of *planning* and *framing assessment* are identified as *before the assessment*. Using assessment data contains two steps—*assessment decisions*, which are identified as *within the assessment protocol,* and *assessment actions,* which are identified as *beyond the assessment protocol but within the event*. Moreover, this framework highlights the importance of looking at teachers' assessment actions after giving feedback, as the learning gaps will not be closed if teachers' feedback is not actually used by students (Purpura & Turner, 2018). This framework provides a perspective on teachers' assessment practices through the lens of the timing of assessment planning and conduct.

In addition, Purpura and Turner (2018) describe different types of teachers' assessment practices helpful in understanding and identifying different forms of assessment activities. In their framework, assessment practices include assessment that is both external and internal to the classroom. *Assessment external to the classroom* is normally carried out as proficiency test, diagnostic or a qualifications exam, essentially a summative assessment. *Assessments internal to the classroom* can include tests and planned assessments, which are usually embedded in teaching, and *spontaneous assessments*, carried out in two forms: *assessment embedded in instruction* and *assessment embedded in conversation* (see Figure 2.7).

Figure 2.7

Purpura and Turners' Model of Assessment Types
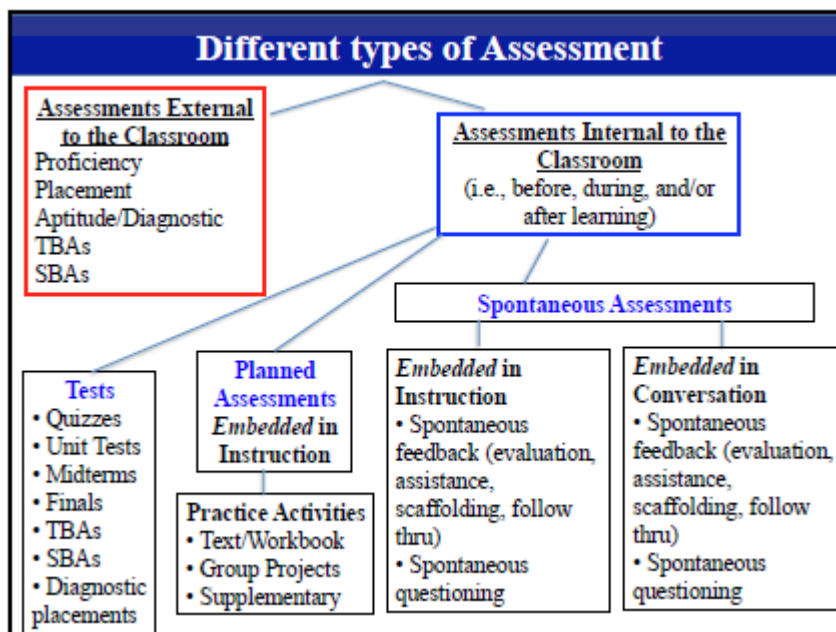
Note. Adapted from Purpura and Turner (2018, LTRC conference)

Figure 2.7 shows that language teachers' assessment practices include summative assessments (tests), planned assessments and spontaneous assessments. As Purpura and Turner point out,

teachers' spontaneous assessments are usually embedded in instruction/teaching or conversations and thus these assessment activities are most often unplanned. Purpura and Turners' two frameworks (Figures 2.6 and 2.7) explore both processes and types of language teachers' assessment practices, expanding understandings of how teachers use spontaneous assessments in classrooms. However, planned and spontaneous assessments are introduced individually in their framework of assessment types (see Figure 2.7), the relationship between these two types of assessments are not mentioned.

Frameworks Featuring Different Assessment Purposes

Three frameworks regarding different assessment purposes were also found in the literature. Cowie and Bell (1999) look into science teachers' formative assessment practices and propose two related forms of formative assessment: planned formative assessment and interactive formative assessment, Furthermore, Harlen (2007) identifies both summative and formative assessment purposes, then proposes a framework of classroom assessment, in which assessments for both purposes are interwoven to improve learning and teaching. Finally, in the language realm, Carless's (2012) LOA framework explores both formative and summative purposes of assessments which integrate in learning tasks in language classrooms.

Cowie and Bell's Framework of Formative Assessment. In the New Zealand science education context, Cowie and Bell (1999) explain the characteristics and functions of formative assessment through a proposed model of formative assessment (see Figure 2.8).

Figure 2.8

Cowie and Bell's Framework of Formative Assessment
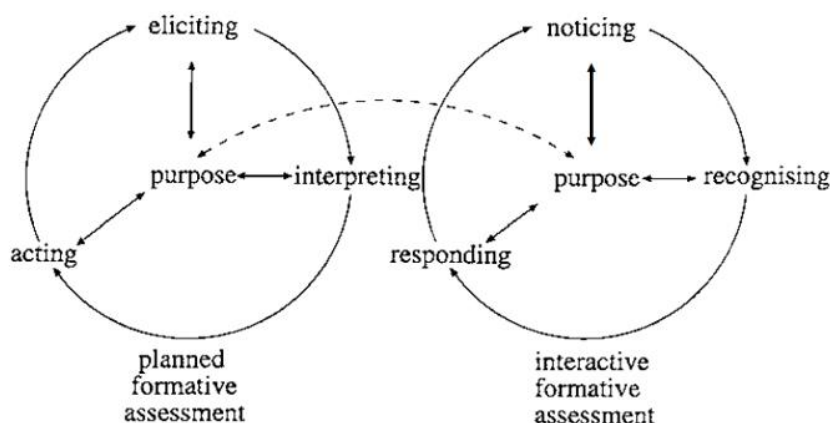
Note. Adapted from Cowie and Bell (1999, p. 113)

By tracing 10 science teachers' assessment practices over 2 years in New Zealand secondary schools, Cowie and Bell identified two forms of formative assessment: *planned formative*

assessment and *interactive formative assessment.* Planned formative assessment describes "how the information was collected, interpreted and acted upon" (p. 103). Teachers gain information about progression of student learning and further adjust teaching accordingly. This kind of formative assessment includes three steps: eliciting, interpreting and acting. Eliciting refers to teachers using planned assessment activities to elicit student knowledge and skills. Interpreting refers to teachers' inference about the extent to which students have learned about the content/skills. After teachers interpret assessment information, they act upon assessment data to enhance student learning. Cowie and Bell suggest three types of feedback from teachers' actions: science-referenced (subject), student-referenced and care-referenced. Regarding interactive formative assessment, Cowie and Bell stress that it takes place in student–teacher interactions. This type of formative assessment usually takes place in learning activities, teachers cannot plan and predict when it occurs. The purpose of interactive formative assessment is to mediate individual student learning. Noticing, recognising and responding are three steps of the process of interactive formative assessment. During interactive formative assessment, teachers notice and recognise temporary but significant information about science, social and personal development from students during or after teaching and learning activities, and then provide immediate responses to students. Unlike planned formative assessment, in which teachers take the initiative to plan and use assessment data, the degree of awareness that teachers have about formative assessment influences whether they can make use of assessment data gained from interactive formative assessment. Cowie and Bell argue that although the distinctive difference between these two types involves different workloads for teachers, both forms show the complex nature of formative assessment and the importance of integrating formative assessment into teaching and learning. Teachers who lack awareness of formative assessment may experience difficulties in identifying different types of assessment.

Cowie and Bell's framework of formative assessment provides valuable insights into the process of interactive formative assessments in the classroom. Furthermore, the interaction between planned and interactive formative assessment in their described framework provides evidence of Black and Wiliam's (2018) claim that the purpose of assessment becomes evident when assessment data is used.

Harlen's Frameworks of Summative Assessment and Classroom Assessment. In the context of the UK, Harlen (2007) investigated assessment practices of science teachers and

proposed three frameworks for teachers' assessment practices: assessment system, summative assessment practices, and assessment processes.

Her framework for summative assessment describes how teachers elicit and use summative assessment evidence (see Figure 2.9). Harlen suggests that unlike formative assessment, which is essentially a repeated cycle in which students can use feedback to improve learning continuously, summative assessment is designed to use the same criteria to evaluate students' performances for grades or levels, and thus the results are not mainly used to indicate future learning and teaching. Harlen's framework provides the major steps of collecting and using summative assessment evidence, nevertheless more assessment purposes and corresponding processes are not discussed.

Figure 2.9

Harlen's Framework of Summative Assessment
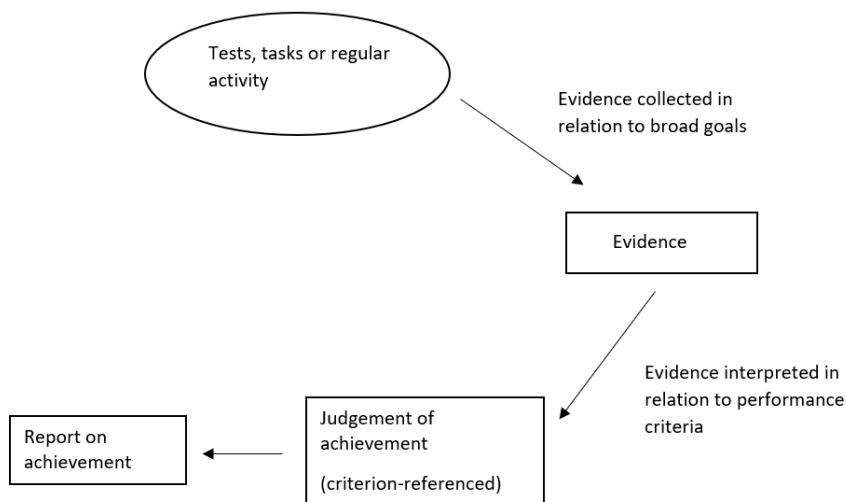
Note. Adapted from Harlen (2007, p. 122)

For a comprehensive understanding of classroom assessment, Harlen developed a model to explore how formative and summative assessments are interwoven in classroom activities (see Figure 2.10). As he explained that there is no clear line between formative and summative assessment; assessment activities often serve both formative and summative purposes. Examples include mock tests which are a kind of "formal formative" assessment. In this model, teachers achieve lesson goals by implementing several learning activities—a, b, c, d etc. Formative assessment is embedded in these learning activities in order to collect evidence of student learning. With this evidence, teachers can improve their teaching and student learning. Moreover, formative assessment evidence can also be used to serve summative purposes after moderation. Thus, both formative and summative assessment purposes can be

realised through classroom formative assessment activities. This model presents the complex process of classroom assessment and sheds light on the possible complementary relationship between teachers' formative and summative assessment practice. However, as this model focuses on discussing the relationship between formative and summative assessments, teachers' overall assessment procedure is not established. Moreover, as this model was developed in science classes, learning activities might be different in second-language classes.

Figure 2.10

Harlen's Model of Learning Activities for Both Formative and Summative Purposes
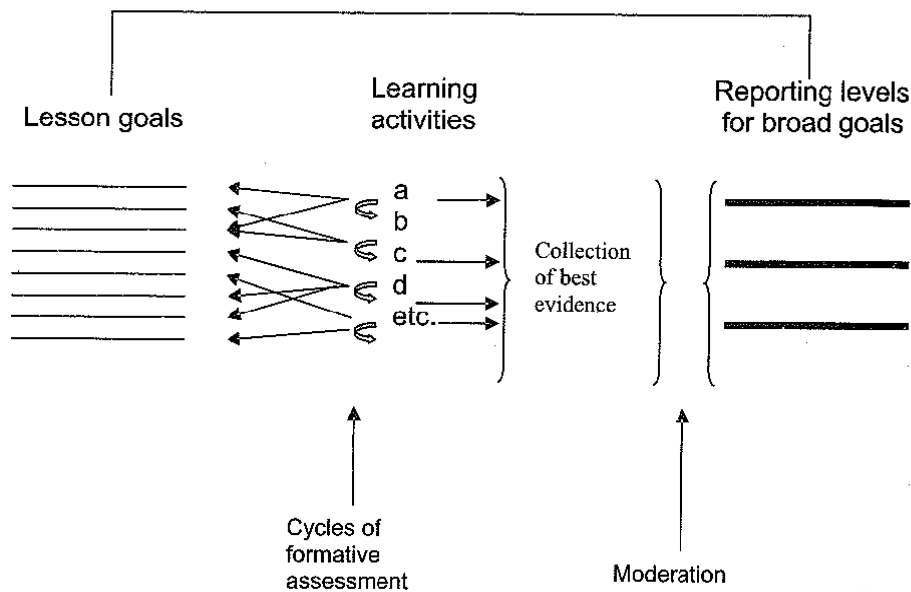
Note. Adapted from Harlen (2007, p. 128)

Carless' Framework of Learning-Oriented Assessment. To effectively incorporate teacher assessment into student learning, Carless (2007) proposes a LOA framework in the Hong Kong tertiary education context (see Figure 2.11).

Figure 2.11

Carless' Framework of Learning-Oriented Assessment
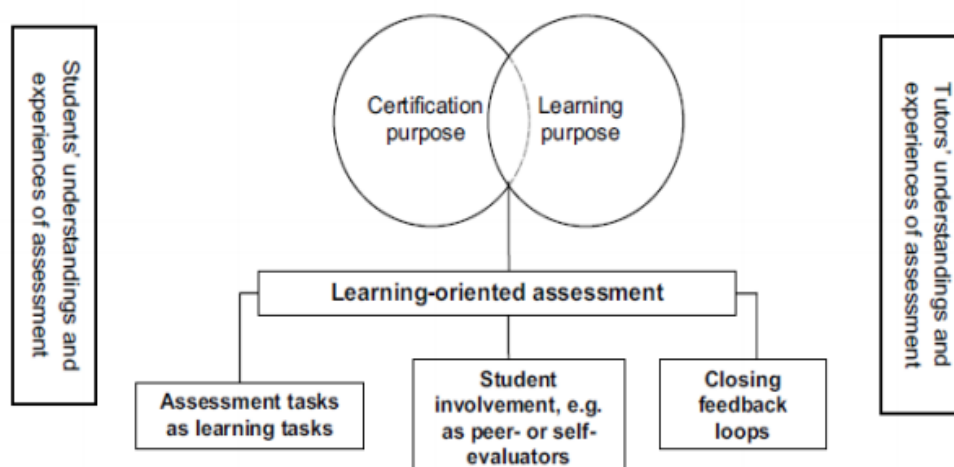
According to Carless, the LOA framework proposes three key features when formative and summative assessment is used mutually to promote learning. The first feature is that assessment tasks are used as learning tasks. Carless mentions that when assessment tasks reflect the objectives of learning tasks as well as the application of language features in the real world, students can use assessment tasks to promote learning progressively. Second, student involvement in assessment tasks is critical to the application of assessment information. Carless suggests that student involvement in assessment tasks could deepen their understandings of assessment standards and learning goals. The third feature is that feedback should be used as feedforward. Carless asserts that feedback does not necessarily link to students' improvement, thus it is critical to engage students in acting on feedback in a timely way.

Carless's LOA framework thus incorporates three key features for language teachers to effectively integrate both formative and summative assessment into student learning. However, Carless's framework of LOA does not explore how teachers employ both planned and unplanned formative assessment activities to promote student learning. This may be because the LOA framework was proposed at the tertiary level in the Hong Kong context, in which influence from summative assessment is dominant. Therefore, informal formative assessment approaches may only be adopted by teachers to a limited extent (Carless, 2007). In contrast to this context, New Zealand has embraced an assessment culture centred around

formative assessment practice. Therefore, there are more possibilities for teachers to employ both planned and unplanned formative assessment in classrooms than this model suggests.

The Conceptual Framework of Assessment Practices for this Study

In this study, a conceptual framework was constructed as a vehicle to identify and understand beginning Chinese-heritage language teachers' assessment practices (See Figure 2.12). The conceptual framework was used as a descriptive tool during data collection. In other words, the conceptual framework was constructed to assist in the identification and description of the beginning Chinese-heritage language teachers' assessment processes rather than to examine the effectiveness of their assessment practices.

Figure 2.12

Conceptual Framework for the Purposes and Processes of Assessment for this Study
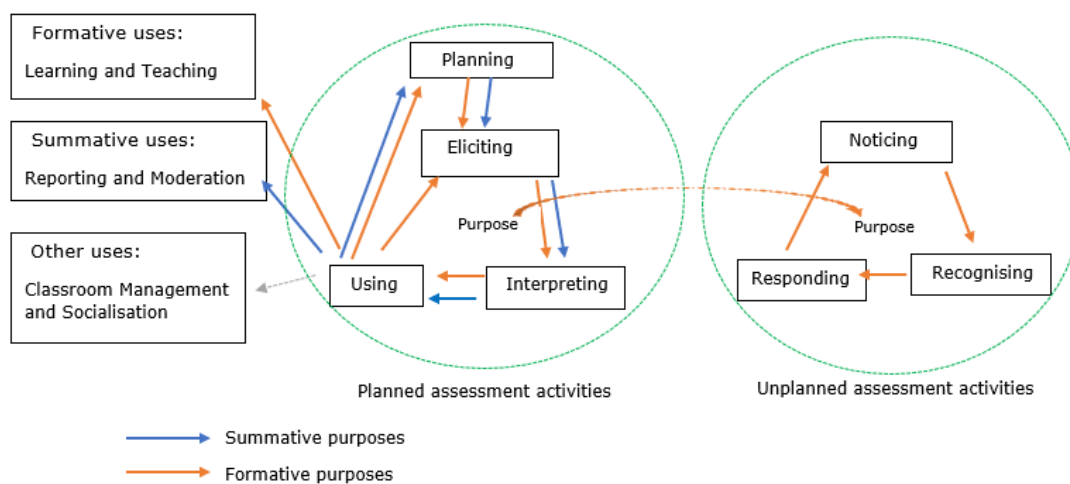
Note. Adapted frameworks from K. Hill and McNamara (2017), Cowie and Bell (1999), and Harlen (2007).

The conceptual framework for the study in this thesis is derived from the reviewed frameworks from K. Hill and McNamara (2017), Cowie and Bell (1999), and Harlen (2007). These frameworks demonstrate the four fundamental steps of planned assessment processes: planning, eliciting, interpreting and using assessment data. Furthermore, the three frameworks cover different aspects of assessment: Bell and Cowie's (1999) frameworks integrate the unplanned formative assessment process with planned forms of assessment, illustrating the interactive relationship between planned and unplanned formative assessment; Harlen's (2007) framework explores the processes of summative assessment; K. Hill and McNamara's (2017) framework suggests various purposes for which teachers may act on assessment information.

Therefore, elements in these three frameworks were incorporated into the one used in this thesis to describe teachers' assessment practices from different aspects.

The described conceptual framework for the study incorporates three aspects of teachers' classroom assessment practice: assessment forms (planned and unplanned assessment activities), processes (stages of teacher assessment practices) and purposes (formative and summative assessment). First, from the aspect of assessment forms, the conceptual framework constructed for this study includes both planned and unplanned forms of classroom assessment. Cowie and Bell's (1999) model of planned formative assessment and Harlen's (2007) model of summative assessment were adapted into the planned assessment aspects of the conceptual framework to indicate that both formative and summative purposes of assessment activities co-exist within classroom assessment (Davison & Leung, 2009; Harlen, 2007; K. Hill & McNamara, 2012). Furthermore, Cowie and Bell's (1999) interactive formative assessment was included as unplanned assessment within the conceptual framework.

From the aspect of assessment purposes, this model integrates both formative and summative assessment purposes into the conceptual framework. This is important because it is not really possible to distinguish between the purposes or processes of formative and summative assessment from observations (Black & Wiliam, 2018; Harlen, 2007). In Figure 2.12, blue arrows indicate summative assessment purposes and orange arrows indicate formative assessment purposes. *Planning*, *eliciting*, and *interpreting* are three shared components for both purposes of assessment. When it comes to the component of *using* assessment evidence, formative and summative assessment purposes become more visible as divergent processes. A formative purpose may be indicated by eliciting evidence or planning to use assessment information in the next cycle (Brown, 2003; Cowie & Bell, 1999; Davison & Leung, 2009), while having a summative purpose leads to assessment evidence being used for *reporting* and *moderation* (Harlen, 2007). Furthermore, the assessment purposes of *classroom management* and *socialisation*, which were illustrated in K. Hill and McNamara's (2012) framework, fit within the *other uses* box in this conceptual framework as these uses are beyond the accountability and teaching/learning purposes of assessment, despite still playing a role in the broader scope of the teacher's classroom activities. The component of *other uses* was included in this framework to explore assessment uses other than formative and summative uses.

Finally, from the aspect of the assessment process, three critical actions in each of the planned (eliciting, interpreting and using) and unplanned (noticing, recognising, and responding) activities were identified from reviewed frameworks above as well as previous literature (Brown, 2002; Cowie & Bell, 1999; Al Sawafi, 2014), and adopted as processes within the proposed framework.

The described conceptual framework (see Figure 2.12) was used to identify and describe assessment processes and purposes in the study reported in this thesis. As described next, in Chapter 3, the framework was used as a reference rather than an observation scheme. Its use, usefulness and limitations are discussed in Chapter 7.

Summary

The broad literature regarding assessment was reviewed in this chapter. In particular, a shift of understanding from summative assessment to integrating both formative and summative assessment, to serve learning, was introduced to provide the rationale for cultivating teachers' assessment capabilities from preservice stage to in-service stage. However, findings of previous literature show that both preservice and beginning teachers face the challenge of understanding and using assessment in practice, their assessment learning is influenced by factors from the personal level, micro level and macro level. In addition, although previous studies have explored the process and different purposes of teacher assessment, observing teachers' complex assessment practices from different aspects remains unclear. Thus, a conceptual framework for observing and understanding teachers' assessment practice from three different aspects is described in the present study.

This study aims to extend our knowledge of beginning Chinese-heritage language teachers' assessment experience in the New Zealand secondary school context, the research question driving the study was: *What are beginning Chinese-heritage language teachers' conceptions and practices of assessment at the completion of their 1-year initial teacher education programme and as they begin teaching over the first 2 years?*

The issues and arguments appearing in the review of the previous literature provide the rationale for this study and further shape the study design which is introduced in the next chapter.

Chapter 3 Research Methodology

Introduction

This chapter first provides a rationale for the adoption of the interpretive paradigm, qualitative methodology, and case-study approach, respectively. The next section revisits the overarching question of this study and presents the six subquestions investigated in two successive phases. The research methods including data collection and data analysis are then described. Finally, I discuss the trustworthiness and the ethical considerations of the study. As noted in Chapter 2, this thesis aimed to explore and describe Chinese-heritage language teachers' conceptions and practices of assessment at the different stages as they move toward full teacher registration in the New Zealand secondary school context. The overarching question for the study was:

> *What are beginning Chinese-heritage language teachers' conceptions and practices of assessment at the completion of the 1-year initial teacher education programme and as they begin teaching over the first 2 years?*

Research Paradigm

This study took place within the interpretive paradigm to investigate how beginning Chinese-heritage language teachers learn about, and use, assessment in the New Zealand secondary school context. Mackenzie and Knipe (2006) indicate that the essence of deciding which paradigm a research study fits within is to "set down the intent, motivation, and expectations for the research" (p. 2). Braun and Clarke (2013) define the term paradigm as "the beliefs, assumptions, values and practices shared by a research community" (p. 4). Similarly, Creswell and Poth (2018) outline the term paradigm as the "worldviews" (p. 18) that researchers carry with them.

In terms of the interpretive paradigm, the fundamental belief is that "reality is socially constructed" (Willis et al., 2007, p. 95). The purpose of research of this kind is to understand a phenomenon in a particular context. Thanh and Thanh (2015) provide an understanding of studies within the interpretive paradigm, stating, "the investigator who follows the interpretive paradigm uses those experiences to construct and interpret his [*sic*] understanding from gathered data" (p. 24). Candy (1989) argues that researchers represent rather than reinterpret events from participants' views. Based on previous research, Phothongsunan (2015) describes three main characteristics of interpretive research: "Interpretive research is conducted with open-ended questions; it focuses on seeking meanings from limited data; and the research results are not for generalisation" (p. 2).

Moreover, as research conducted within the interpretive paradigm values human behaviour, research in this paradigm is more likely subjective (Thanh & Thanh, 2015; Willis et al., 2007). In the present study, beginning Chinese-heritage language teachers' prior and current conceptions and practices of assessment are valued and used to understand their professional growth in assessment. Angen (2000) argues that it is worth taking the risk with the validity and reliability of research in using interpretive research, for it deepens "understanding of how humans experience the lifeworld through language, local and historical situations, and the intersubjective actions of the people involved" (p. 386).

Methodology

An interpretive paradigm is usually associated with qualitative research methodology. Mackenzie and Knipe (2006) compared the definitions of methodology and method, pointing out that methodology is "the overall approach to research linked to the paradigm or theoretical framework while the method refers to systematic modes, procedures or tools used for collection and analysis of data" (p. 5). Denzin and Lincoln (2017) consider qualitative research as the attempt to investigate "a situated activity that locates the observer in the world, it consists of a set of interpretive, material practices that make the world visible" (p. 43). Qualitative methodology is usually used to uncover the similarities, contrasts and even conflicts between behaviours, beliefs, opinions, emotions, and relationships of individuals (Mack et al., 2005).

Braun and Clarke (2013) define qualitative research by summarising the fundamental characteristics that are different from quantitative research. In their definition, qualitative research has the focus on seeking meaning rather than numbers; the context of this type of research is usually highlighted. In this study, a qualitative methodology was helpful to unpack the assessment experience of the group of Chinese-heritage language teachers in the New Zealand context, identifying mediating factors that have influenced their changing conceptions and practices of assessment over time. Methods within qualitative methodology, such as interviews, classroom observations, and reflective conversations were adopted to obtain and present participants' experience of learning about, and using, assessment.

Case-Study Approach

As stated, this study set out to document and understand the phenomenon of Chinese-heritage language teachers' assessment conceptions and practices as they began teaching in the New

Zealand secondary school context. To answer the research questions, a multiple case-study approach was employed to understand their different stories of assessment learning.

Previous research has identified divergent perspectives on the case-study approach in social sciences (Creswell, 2007; Denzin & Lincoln, 2017; Merriam, 1998; Yazan, 2015). Schwandt and Gates (2017) indicate that a critical problem for researchers who employ a case-study approach in their research is its varying definitions. Yazan (2015) reviews previous research and summarises two main different perspectives on understanding case studies. Yin's (2014) perspective, underpinned by positivist philosophy, establishes that the purpose of conducting case studies is to "retain a holistic and real-world perspective" (p. 4). Based on this understanding, Yin (2002) identifies "case" as "a contemporary phenomenon within its real life context, especially when the boundaries between a phenomenon and context are not clear and the researcher has little control over the phenomenon and context" (p. 13), and thus case study is used to inquire into the phenomenon and address "how" and "why" questions.

From a different perspective, Stake (1995) argues that social facts are interpreted and presented by researchers in case studies, and thus there is no best view to understand cases (p. 108). Drawing on the philosophy underlying constructivism, Stake (1995) identifies the term *case* as "an integrated system which has a boundary and working parts," it is "an object rather than a process" (p. 2). Therefore, case study is used, in this view, to understand the particularity and complexity of the case within a specific context (Stake, 1995). Merriam (1998) broadens Stake's definition by referring to the function of case study as describing and analysing any bounded system; researchers could identify what a case is, as long as they are able to clarify research objectives. She indicates that the purpose of case studies is to understand stories from participants' perspectives, and thus findings of case studies are co-constructed by both researchers and participants In this interpretive case study, underpinned by constructivist philosophy, the primary goal is to uncover participants' knowledge and practices, and how they come to be constructed in that way (Yazan, 2015).

Nevertheless, some common features of case studies among these different understandings can be established. Firstly, the case-study approach is employed to explore the interrelationship between phenomenon and context. Furthermore, case study sits in the interpretive (sometimes referred to as constructivist) paradigm that highlights that reality is socially constructed and case studies yield rich descriptive data. In other words, participants in case studies are encouraged to describe and explain reality from their perspectives, and

researchers explore the phenomenon through participants' responses. Also, case studies aim to inquire into either a bounded system (a case) or multiple bounded systems (cases), and thus methods such as interviews, observations, and concrete materials can be used to collect detailed and in-depth data (Baxter & Jack, 2008; Creswell, 2007; Merriam, 1998; Simons, 2009; Stake, 1995).

The investigation for this thesis adopted a case-study approach for the following three reasons. Firstly, the study aimed to uncover Chinese-heritage language teachers' conceptions and practices of assessment as they began to teach in New Zealand secondary schools. The case-study approach was useful to explore the interrelationship between the New Zealand secondary school context and beginning Chinese-heritage language teachers' conceptions and practices of assessment.

Secondly, as the present study intended to explore and understand participants' experience of assessment from their perspectives, the case-study approach was considered appropriate, as "researchers cannot manipulate the behaviour of those involved in the study" (Baxter & Jack, 2008, p. 545).

Finally, a major advantage of adopting a case-study approach is that researchers employ multiple methods to approach the phenomenon of interest. As stated above, methods of semi-structured interviews, classroom observations, reflective conversations, and documentary data were employed in this study to collect rich descriptive and in-depth data about how participants conceptualised and implemented their assessment practices across the 2-year period.

In the present study, a multiple case-study approach was employed to explore the similarities and differences in individual participants' experience of assessment learning. Creswell (2007) distinguishes multiple case studies from single case studies by identifying research purposes: "Often the inquirer purposefully selects multiple case studies to show different perspectives on the issue" (p. 74). Furthermore, the multiple case-study approach is usually employed to understand the similarities and differences between cases (Baxter & Jack, 2008; Gustafsson, 2017; Stake, 1995). Also, according to Baxter and Jack (2008), multiple case studies allow researchers to explore the influences of contextual factors on individual cases as well as across cases. Based on these understandings, multiple methods (interview, classroom observation, reflective conversation, document analysis) were employed to:

1) Compare individual teacher's conceptions of assessment at different phases (e.g., teachers' conceptions of assessment when completing the ITE programme and during their first 2 years of teaching);

2) Compare different teachers' conceptions and practices of assessment in the same phase;

3) Compare the influences of different mediating factors, such as Chinese heritage and examination systems in different contexts.

A typical criticism of case study is that there is no basis for scientific generalisation. However, Stake (1995) argues "only a single case or just a few cases will be studied at length. Certain activities or problems or responses will come up again and again. Thus, certain generalizations will be drawn" (p. 7). Moreover, as noted by Yin (2014), scientific generalisation includes two types: statistical generalisation and analytic generalisation. It is difficult if not impossible to realise statistical generalisation in qualitative case studies, for "the small numbers of cases in case studies are hard to apply findings in a large population" (p. 68). In contrast, case studies are more likely to realise analytic generalisation, because "findings of case studies are used to implicate other conditions" (p. 68). In the present study, the researcher repeatedly gathered and observed participants' conceptions and practices of assessment at different stages as they began to teach. Therefore, their long-lasting conceptions, as well as changing conceptions surrounding assessment, were investigated. Their stories of assessment learning emerged and shed empirical light on some theoretical concepts in relation to beginning Chinese-heritage language teachers' conceptions of assessment.

Research Questions and Overall Study Design

A longitudinal dimension was employed in this study to investigate beginning Chinese-heritage language teachers' conceptions and practices of assessment from the preservice stage to the in-service stage. The longitudinal dimension increases the validity and reliability of the study, for participants had more space to express themselves and explain their conceptions and practice of assessment as they began to teach. Based on this, the investigation comprised two successive phases. Guided by the research questions, Phase 1 investigated preservice Chinese-heritage language teachers' assessment conceptions and how well they believed they were prepared for using assessment during the 1-year initial teacher education (ITE) programme. The guiding question in Phase 1 was:

> *What are preservice Chinese-heritage language teachers' conceptions of assessment at the completion of their 1-year initial teacher education programme?*

Three subquestions were used for shaping Phase 1:

4) What are preservice Chinese-heritage language teachers' conceptions of assessment at the completion of their 1-year initial teacher education programme?

5) What are the factors influencing preservice Chinese-heritage language teachers' conceptions of assessment?

6) How well prepared do they feel to put their assessment understandings and skills of assessment into classroom practices?

The Phase 2 study then examined the beginning Chinese-heritage language teachers' conceptions and practices of assessment after they had started to teach in New Zealand secondary schools. The guiding question in the Phase 2 study was:

What are beginning Chinese-heritage language teachers' assessment conceptions and practices and how do these change as they progress through their first 2 years of teaching?

The following subquestions guided Phase 2:

4) What are beginning Chinese-heritage language teachers' conceptions of assessment as they move towards full teacher registration?

5) What are beginning Chinese-heritage language teachers' practices of assessment as they move towards full teacher registration?

6) What factors influence beginning Chinese-heritage language teachers' assessment conceptions and practices as they begin teaching?

The overview of the study design is illustrated in Figure 3.1.

Figure 3.1

Overview of the Study Design

Research question:
What are beginning Chinese-heritage language teachers' conceptions and practices of assessment at the completion of their 1-year initial teacher education programme and as they begin teaching over the first 2 years?

Paradigm: Interpretivist

Methodology: Qualitative study

Case study approach

Data collection

Phase 1: Preservice teachers' conceptions of assessment

Phase 2: Beginning teachers' conceptions and practices of assessment

The focus group interviews

Semi-structured interviews

Classroom observations

Reflective conversations

Documents

Reflexive thematic analysis
Inductive approach

Data Collection

As indicated above, this study consisted of two phases of investigation. The Phase 1 study investigated beginning teachers' conceptions of assessment at the completion of their 1-year initial teacher education (ITE) programme, while the Phase 2 study explored the assessment conceptions and practices of assessment of three participants at different stages over a 2-year period as they moved towards full teacher registration. Pseudonyms were used for all the

participants throughout the study. The overview of the data-collection procedure throughout the study is outlined in Figure 3.2.

Figure 3.2

The Procedure of Data Collection

Phase one: Preservice Chinese-heritage language teachers' conceptions of assessment

> One focus group interview at the end of 2016:
>
> Seven preservice Chinese-heritage language teachers

Three preservice teachers volunteered to continue to participate in phase 2

Phase two: Beginning Chinese-heritage language teachers' conceptions and practices of assessment

Participants	Stage 1	stage 2	stage 3	stage 4
	Amber, Diana and Michelle	Amber, Diana and Michelle	Amber and Michelle	Amber and Michelle
	December 2017	June 2018	September 2018	November 2018	**Dates**

Methods: semi-structured interviews, classroom observations and reflective conversations

In Phase 1, a focus group interview was used to probe into research questions regarding preservice Chinese-heritage language teachers' conceptions of assessment before they began teaching. Seven participants shared their experience of learning assessment in the initial teacher education programme.

Phase 2 was carried out across the first two years as the beginning Chinese-heritage language teachers moved toward becoming fully registered teachers. Thus, Phase 2 consisted of four stages of data collection, undertaken in Term 4 of 2017 and in each term during the academic year of 2018 (see Figure 3.2). Interviews, classroom observations, reflective conversations and documents were collected and triangulated regarding participants' conceptions and

practices of assessment. Reflexive thematic analysis (Braun & Clarke, 2013, 2019) was adopted to explore data in depth in both phases.

The following subsections provide more detailed descriptions regarding settings, data-collection methods and procedures of each phase of data collection.

Phase 1

The Phase 1 study was carried out in a university setting to explore preservice Chinese-heritage language teachers' conceptions of assessment at the completion of their 1-year teacher education programme in New Zealand. Preservice language teachers with Chinese-heritage backgrounds were recruited to share their assessment learning stories at the completion of their 1-year initial teacher education (ITE) programme.

Participants. In Phase 1, purposive sampling was employed to recruit suitable participants. The purposive sampling strategy is used to select particular participants, settings, events or processes in order to provide a detailed understanding of the research problem (Maxwell, 2012; Palys, 2008). Nevertheless, research employing this strategy to select participants may not represent any larger population (Maxwell, 2012; Yin, 2014).

Participants who met the following three criteria were invited. Firstly, they were invited if they planned to teach Chinese language in New Zealand secondary schools. In the New Zealand context, secondary school teachers are responsible for using assessment to both promote students' learning and determine their academic achievement. Secondly, participants were preservice Chinese-heritage teachers who had received at least 9 years of compulsory education in the Chinese context. As one purpose of this study was to probe into Chinese-heritage language teachers' conceptions and practices of assessment as they shifted from the Chinese context to the New Zealand context, this criterion was employed to ensure that participants had strong backgrounds in Chinese culture. Thirdly, participants were invited if they had studied the course "Teaching Languages" in the ITE programme.

At the time when ethics approval was gained (September 2016), a cohort of nine preservice Chinese language teachers was in the last semester of the ITE programme. With permission from the programme leader (see Appendix A and Appendix B), and the lecturer of the course Teaching Languages, this cohort of preservice teachers was invited to participate during their class break. Seven preservice language teachers volunteered as participants for the focus group interview. All of them received participant information sheets (see Appendix C) and

signed consent forms (see Appendix D) before the focus group interview was conducted in November 2016.

These participants shared some common features. Firstly, all the participants were from the same class and were familiar with each other. Thus, they were likely to discuss topics in a natural way (Zhou, 2016). Moreover, all of them were female, with ages varying from 20 to 40. In addition, none of them had received any training regarding assessment in the New Zealand context prior to the initial teacher education programme. Nevertheless, three participants demonstrated that they had had rich teaching experience in China (8 years, 8 years and 11 years respectively) with no teaching experience in New Zealand; three participants had schooling experience in New Zealand (4 years and 8 years respectively) while four had no schooling experience in the New Zealand context. However, one participant had 1 year of teaching experience in New Zealand primary and secondary schools as a Mandarin language assistant.[3] Table 3.1 summarises information about Phase 1 participants' schooling and teaching experience in both China and New Zealand.

Table 3.2

Participants' Schooling and Teaching Experience Prior to the Initial Teacher Education Programme

Name	School experience in China	School experience in New Zealand	Teaching experience in China	Language-teaching experience in New Zealand
Michelle	16 years	-	-	1 year
Emma	16 years	-	8 years	1 month
Ava	16 years	-	8 years	3 years
Chloe	11 years	7 years	-	-
Diana	17 years	-	11 years	-
Amber	10 years	7 years	-	-
Mia	13 years	4 years	-	-

---

[3] Mandarin language assistants (MLAs): The Chinese government-funded Mandarin language assistant programme is open to primary, intermediate, secondary or tertiary institutions in New Zealand for new or existing Mandarin programmes. Kai (2015) explains the MLA's role is to raise the cultural awareness of both teachers and students in local schools.

Setting: The Initial Teacher Education (ITE) Programme. Participants in this study were from the same university, located in a large city in New Zealand. Programmes in teacher education, social work, and some other programmes such as physical education were provided in this faculty. All the participants of this study had completed their bachelor's degrees (some had completed master's degrees) and enrolled in an ITE programme, the Graduate Diploma in Teaching (Secondary). It is an intensive 1-year ITE programme which enables students to earn a teaching qualification required as the first part of becoming a fully registered teacher in order to teach in New Zealand secondary schools. Aligned with the expectation of New Zealand's education system, this ITE programme aims to develop "research-informed, inquiry-based practitioners who lead the learning of others by creating purposeful, responsive and safe learning environments within which all learners achieve valued skills, knowledge and attitudes" (Initial Teacher Education Programme Handbook, 2006).

In the academic year of 2016, around 200 preservice secondary teachers enrolled in this ITE programme. They were required to complete the two core elements of the programme, Te Whakaako in NZ Secondary Schools and Professional Learning in Practice, and selected courses regarding teaching specific subjects connected to their interests. Te Whakaako in NZ Secondary Schools is a course that focused on introducing the *New Zealand Curriculum*, pedagogy and evidence-based assessment. preservice teachers were expected to develop their understandings and knowledge to cater for student motivation and engagement in New Zealand secondary schools. Furthermore, preservice teachers were required to complete the core element of this programme, Professional Learning in Practice, which included the two practica (practice placements) to develop their knowledge and skills for effective teaching and assessment. In addition, within each course, this programme provided lectures, tutorials and workshops for students to deepen their understanding and motivate them to reflect on their own teaching and assessment practice. Finally, preservice teachers in this programme completed 30 weeks of preservice courses and 14 weeks practica in secondary school classrooms (see Figure 3.3).

Figure 3.3

Timetable of the Initial Teacher Education Programme

																Teaching Week													
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

Semester 1: March to July | Semester 2: July to November

Teaching placement (Practicum) in school (different school each semester)
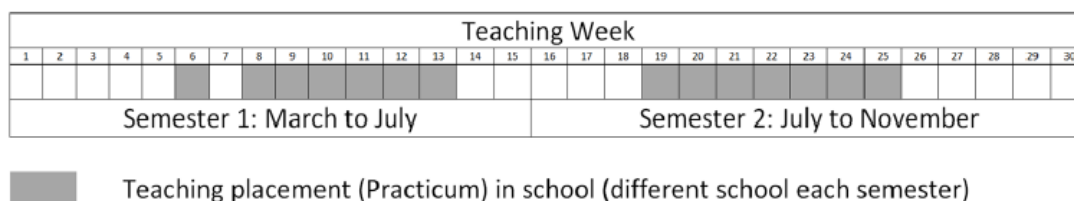
Note. From East (2016b, p. 9).

Preservice teachers were engaged in coursework in the first 5 weeks in Semester 1; they then observed language classes in a secondary school classroom for a week (Week 6). On their return to campus in Week 7, the lecturer guided them to discuss and set up effective teaching plans before they undertook the rest of Practicum 1 (Weeks 8–13) in assigned schools. Finally, after student teachers finished the first practicum, they had 2 weeks of reflective learning in the university setting. Semester 2 had a similar arrangement. East (2016b) indicates that the cyclical nature of this programme allowed preservice teachers to reflect upon their practices and develop their capabilities as reflective practitioners.

All participants in this project had completed the three courses in the ITE programme: Te Whakaako in NZ Secondary Schools (Teaching in NZ Secondary Schools), Teaching Languages, and Teaching the Chinese Language. In line with the requirement of the Graduating Teacher Standards (New Zealand Teaching Council, 2007), the content of these courses covered both the principles of the *New Zealand Curriculum* for teaching languages and practical skills of planning and delivering teaching and assessment practices. Therefore, preservice teachers had had the opportunity to integrate knowledge and skills into their teaching practice to "maximise student motivation, engagement and success in learning the target language" (New Zealand Teaching Council, 2007).

As stated above, every preservice teacher was also required to undertake two practica in the ITE programme (see Figure 3.3). The purpose was to provide opportunities for preservice teachers to develop research-informed, inquiry-based practice in authentic settings (Initial Teacher Education Programme Handbook, 2016). Prior to the focus group interviews in Phase 1, all the participants had undertaken practica in two different secondary schools. All observed experienced teachers teaching, although some participants reported that they also had opportunities to teach and take part in assessment in their practicum school. As a result, their experience of practica varied.

After the focus group interview at the end of 2016, all the participants graduated from the ITE programme and successfully gained provisional registration from the New Zealand Teaching Council. While some of the participants started to look for teaching positions in New Zealand, some of them had already received offers to teach Chinese language in secondary schools.

Method: Focus Group Interview. In Phase 1, a focus group interview was carried out in a quiet classroom in the university at the end of the 2016 academic year, when seven participants had completed all courses and prepared for graduation from the ITE programme. They were invited to a 60-minute semi-structured interview, sharing and discussing how the 1-year ITE programme prompted them to understand and use assessment in the New Zealand secondary school context. Their responses in the focus group interview were audio recorded for transcription and analysis.

The research method of focus group interviews was considered appropriate in the Phase 1 study for the following reasons. First, the focus group interviews were identified as a method of data collection to explore insights from group interactions (Ho, 2006; Lederman, 1990; Litosseliti, 2003). Participants in a focus group interview are encouraged and prompted to exchange ideas and anecdotes during the discussion of the interview questions (Ho, 2006; Lederman, 1990). Lederman (1990) argues that focus group interviews are useful to "probe the inner workings of the minds of the group members" (p. 119). In Phase 1, as all the preservice language teachers had Chinese-heritage backgrounds, employing the focus group interview was useful to explore influences from CHCs on participants' conceptions and practices of assessment.

Furthermore, the synergy of ideas within group interactions can stimulate richer data than individual interviews (Lederman, 1990). Lederman (1990) stresses that interactions within the group can generate new ideas regarding interview questions, which normally cannot be found in individual interviews. Ho (2006) also claims that when focus group interviews are well designed, interactions within groups could provide in-depth understandings of the topic under study. In the current study, a set of guiding questions (see Appendix E) were adopted to ensure the focus group discussion was focused on the research questions. Meanwhile, the participants were encouraged to talk freely about their assessment experience and learning in the 1-year ITE programme. The interactions within the focus group flowed naturally from one research question to another. As a moderator, I observed participants' interactions and tried to keep my influence on their discussions to a minimum.

Nevertheless, the focus group interview approach has its limitations. Kitzinger (1995) points out the potential risk that individual voices or disagreement may go unheard during group discussions. Therefore, the approach of individual interviews was adopted in the Phase 2 study to elicit each participant's assessment experience in depth as they began to teach in New Zealand secondary schools.

Language of Data Collection. During the focus group interview, although participants all came with Chinese-heritage backgrounds, they were encouraged to reflect in English on their experience of assessment for the following reasons. Firstly, as the participants had all been immersed in learning and practising assessment knowledge during the 1-year ITE programme in English, they were confident in understanding and answering questions in English. Furthermore, as they were provided opportunities to reflect on their experience of assessment during the reflection sessions after practicum (East, 2016b), it was appropriate for participants to use English to recall prior experience of assessment, especially using professional terminology. In addition, using English language in the focus group interview could reduce misinterpretation and misunderstanding during the process of data analysis.

Phase 2

The aim of Phase 2 was to continue examining beginning Chinese-heritage language teachers' conceptions and practices of assessment as they began teaching in New Zealand secondary schools (see Figure 3.1). Phase 2 was a longitudinal study, including four stages of investigation within the participants' first 2 years as provisionally registered teachers in the New Zealand context. Methods of semi-structured interviews, classroom observations, reflective conversations, and documentary review were employed to elicit data.

After graduation from the ITE programme, all the participants in Phase 1 were successfully granted provisional registration, which allowed them to teach in New Zealand secondary schools. In the New Zealand context, provisionally registered teachers normally take 2 years of teaching within a broad-based programme of induction and mentoring before they can apply for full teacher registration (Cameron, 2007). Based on this understanding, Phase 2 was conducted to investigate participants' conceptions and practices of assessment during the first 2 years as provisionally registered teachers. The invitations to continue to participate in Phase 2 were sent to all seven Phase 1 participants in November 2017 to ensure that, by then, most of participants had found teaching jobs in schools. As a result, three participants from Phase

1, Diana, Amber and Michelle, confirmed their participation for Phase 2 (see more details in Chapter 6).

To initiate the research process, an invitation letter introducing this project was sent to each school principal and head of department for each potential participant. After receiving their agreement, each participant received a participant information sheet (see Appendix F) and a consent form (Appendix G) as well as an introduction about this phase of the study from the researcher. Each stage of the Phase 2 investigation (see Figure 3.2) included one interview, a classroom observation and corresponding reflective conversation after each classroom observation, and collection of any relevant documents. These multiple strategies were helpful to explore and understand each individual participant's assessment experience at the different stages as they moved toward full teaching registration. Table 3.3 provides an overview of the Phase 2 research design as described above.

Table 3.3

Data Collection Overview of Phase 2: Beginning Chinese-Heritage Teachers

Data collection	Data-collection methods	Data-collection timeframes	Descriptions
Beginning Chinese-heritage language teachers' conceptions and practices of assessment as they move towards full teaching registration.	Interviews	Four times between mid-November 2017 and mid-December 2018.	Up to 60 minutes semi-structured interviews conducted with four participants individually.
Beginning Chinese-heritage language teachers' practice of assessment as they move towards full teaching registration.	Classroom observations	Four times between the end of November 2017 and end of November 2018.	Audiotaped 45- or 80-minute language class in individual participant's classroom.
Beginning Chinese-heritage language teachers' conceptions and practice of assessment as they move towards full teaching registration.	Reflective conversations	Four times between the end of November 2017 and end of November 2018.	Up to 20 minutes audiotaped conversations after each classroom observation.
Factors influencing beginning Chinese-heritage language teachers' assessment conceptions and practices as they begin teaching. Data regarding beginning Chinese-heritage language teachers' conceptions and practices is not found in interviews and observations.	Documentary data	Throughout the study.	School curricula guiding language learning; educational assessment policies in schools, teachers' personal records, and assessment sheets or activities that the teachers used.

Methods. As noted above, Phase 2 employed four main methods to collect and triangulate data from different sources. Denzin and Lincoln (2017) explain that qualitative research is the representation of the phenomena in which researchers employ a series of methods to interpret and make sense of meanings that people carry with them. With this in mind, semi-structured interviews, classroom observations, reflective conversations, and document analysis were adopted to describe and understand the complex reality of assessment from participants' perspectives. The following sections describe and justify data-collection methods, as well as procedural information about the use of these methods.

Semi-Structured Interviews. Semi-structured interviews were used as a primary method to collect rich and in-depth data in this phase of the investigation. The interview is a major approach to conducting qualitative research, partly due to the fact that the complexity of people's experience encourages researchers to explore the unknown (Kvale & Brinkmann, 2009). It is a method ideally suited to collecting information from participants about their previous experience and their in-depth perceptions on particular topics (D. W. Turner, 2010).

According to Harrell and Bradley (2009), interviews can be considered on a continuum, in terms of structure, ranging from unstructured interviews, semi-structured interviews, to structured interviews. The major difference between these three types is the extent to which the interviewer dominates the interaction. Wilkinson and Birmingham (2003) provide a rationale for using semi-structured interviews, where the interviewer designs questions with a more flexible structure so that participants can have more space to explore their conceptions. A set of interview questions is designed, along with some ideas for probing the answers for further information where necessary. In this way, the interviewer can collect more in-depth, meaningful information based on interviewees' responses. Given this consideration, both the focus group interview in Phase 1 and individual interviews in Phase 2 were semi-structured.

In Phase 2, each semi-structured interview lasted no longer than 60 minutes and was conducted at a time and place convenient to the beginning teachers, within the proposed timeframes. Interview questions in each stage during the Phase 2 investigation were designed with different foci (see Table 3.4). Furthermore, the interview questions for each beginning teacher in each stage were adapted based on the analysis of their responses from the previous interviews. To give a more specific explanation, the interview questions at Stage 1 stemmed from the three participants' responses during the focus group discussion in the Phase 1 study;

the interview questions at Stage 2 followed from answers at Stage 1, and so on. Participants'
responses were audiotaped after receiving their permission.

Table 3.4

Foci of the Interviews in Phase 2: Beginning Chinese-Heritage Teachers

Stages	Data-collection Timeframes	Focus	Format
Stage 1	End November 2017	Participants' assessment experience in their first year.	Open-ended questions
Stage 2	End February 2018	Participants' experience of assessment at the beginning of second year.	Semi-structured and open-ended questions
Stage 3	Mid-June 2018	Discussion about effective assessment practices and factors contributing to their conceptual changes.	Semi-structured and open-ended questions
Stage 4	Mid-December 2018	Reflections on assessment experience, discussion of the development of assessment and corresponding influential factors.	Semi-structured and open-ended questions

In addition, during individual interviews, a responsive interviewing strategy was employed to
modify interview questions and motivate participants to explore their conceptions. Rubin and
Rubin (2005) argue that the rationale for applying a responsive interview approach in
qualitative research, is that "qualitative interviewing is a dynamic and iterative process, not a
set of tools to be applied mechanically" (p. 15).

Classroom Observations. In Phase 2, classroom observations were employed to record
participants' actual assessment practices. Classroom observation is a qualitative research
method that can "enable [a] researcher [to] discover things participants are not aware of, or
that they are unwilling to discuss" (Hoepfl, 1997, p. 53). During observations, observers are
able to capture first-hand data in natural settings (Cohen et al., 2007; Merriam, 2009; Patton,
2014). Patton (2014) points out that observations provide opportunities to describe the context
where the meanings of what was observed emerge, and thus observers can understand the
context in a holistic manner.

Four observational approaches of "complete participants," "participant-as-observer,"
"observer-as-participant," or "complete observer" are commonly used to distinguish between
approaches in observational inquiries. Observers choose the extent of their participation in the

action to understand their research subjects through verbal and non-verbal communications (Cohen et al, 2007; Patton, 2014; Wilkinson & Birmingham, 2003). Hatch (2002) lists four factors that influence observers' decisions on the extent of participation in observations. First, participation of observers may influence natural settings and cause the context to become "unnatural." Second, the observers' capability as participants will influence the settings. Hatch (2002) points out that, "less evident are problems that educational researchers face when attempting to participate as teachers or administrators" (p. 74) while also observing. The third factor relates to observers' abilities of time management. Hatch (2002) indicates that observation data may be missing when observers both participate in teaching and keep field notes.

In Phase 2, I took the role of a complete observer to capture beginning Chinese-heritage language teachers' classroom assessment practices in natural settings. As pointed out by Shaughnessy et al. (2011), the difficulty of conducting classroom observations as a complete observer is that the researcher's presence in the classroom may lead to participants' reactivity. In other words, teachers and students may change their actions when they know they are being observed. For instance, teachers may increase the number of assessment activities or change methods of teaching content. To minimise such influence, both students and teachers were informed of the observation schedule during my first introduction in the observation classrooms. Furthermore, I extended my presence in classrooms, spending time in the classrooms before the observations took place so that teachers and students could get used to my presence (Shaughnessy et al., 2011).

As a complete observer, I sat as unobtrusively as possible in the classroom. Beginning teachers' classroom activities were audiotaped during observations. It is noted that only the three beginning teachers' practices were observed and recorded during classroom observations, students' actions were not documented because the focus of observations was beginning teachers' classroom assessment practices. To achieve this purpose, the three participants wore portable clip-on microphones during observations to record their instructions and interactions with students in a detailed and accurate way. This also minimised the risk of recording student responses. Audiotaping is an effective strategy to capture the precise wording used by teachers (Creswell, 2011; Dixon, 2008). Audiotaped records were used to triangulate data from interviews and documents.

In addition, to complement audiotaped data, I took field notes during classroom observations. Recording field notes is a practical strategy to collect rich data (Morrell & Carroll, 2010). Based on the conceptual framework for this thesis (see Figure 2.12), field notes recorded both descriptive and reflective information regarding processes (planning, collecting data, interpreting and using assessment data), purposes (formative and summative purposes) and forms (planned and unplanned assessment activities) of participants' classroom assessment practices. Also, information included descriptions of classroom settings, actions of teachers, and reflective information including my ideas, thoughts, and concerns that occurred on site.

In Phase 2, four rounds of classroom observations were conducted at the end of 2017 academic year and during the 2018 academic year, with a total of 12 lessons observed. Diana and Amber were observed in eight double lessons (80 minutes each), and Michelle was observed in four single lessons (50 minutes each). Class sizes for all observation lessons were small and ranged between 10 and 15 students.

Reflective Conversations. After each classroom observation, participants were invited to have reflective conversations, in order to clarify any unclear aspects observed during classroom observations. The purpose of conducting reflective conversations was to complement the observational data. Creswell (2011) indicates that compared with participant observers, complete observers may not be able to gain concrete observational data, as they choose to "remove them themselves from actual experiences" (p. 215). In other words, researchers may miss important information regarding observed objectives, given that complete observers keep distant from what happens in classrooms, particularly interactions between teachers and students. Therefore, reflective conversations took place immediately after classroom observations in empty classrooms or school cafeterias. Heflich and Rice (1999) identify reflective conversation as "an exchange of ideas in which the expression and receipt of ideas leads to the construction of new understanding of their own experience among the participants" (p. 3). They stress that reflective conversation is not only a research tool to inquire into teachers' teaching practice, it also provides an opportunity to discuss "our intuitive knowledge base things that have previously occurred" (p. 3). In the current study, questions were specific to each participant, emerging from the fieldnotes made during observations. During the reflective conversations, participants were first asked about their overall impressions of the observed lessons. By reflecting on the whole procedure of conducting assessment, participants had opportunities to reflect upon how their classroom assessment practice had been successful or otherwise. Building on their overall reflections, more specific questions were asked relating

to unclear parts of lessons during classroom observations. During my inquiry into their reflections, I stopped/paused intentionally to leave them time to refresh their memories. All the reflective conversations were audiorecorded. After having conversations with teachers, I found quiet places nearby and recalled the information I needed from the conversations and then added it into my observation records and field notes.

Document Review. Documents, including public and participants' private records of teaching plans and student performances, were elicited during the data-collection procedure to corroborate data from observations and interviews. Public records included the *New Zealand curriculum*, the assessment policies of each participant's school, school handbooks, guidelines of assessment, each school's annual plan for 2018, and examination schedules. I also collected the course descriptions, test papers, achievement standards and assessment schedules relevant to the interviews and classroom observations from each class (see an example Appendix H). Private records such as participants' notes on exams and student performances were collected (see an example Appendix I). Documents are useful resources for researchers, "not only because of what can be learned directly from them but also as a stimulus for paths of inquiry that can be pursued only through direct observation and interviewing" (Patton, 2002, p. 294) Thus, documentary data in this study was utilised to complement and triangulate the data generated from classroom observations, reflective conversations and interviews, thereby increasing the credibility of the study (Naruemon, 2013). Furthermore, documentary data were helpful to understand participants' conceptions of assessment and the contexts where teachers shape and adjust their assessment practices. According to Lasky (2005), teachers' professional careers are eventually shaped or affected by school culture, school policy and education policy concerning accountability.

Language of Data Collection. To maximise data collection from interviews and reflective conversations and ensure the voluntary nature of the study, the three participants in Phase 2 were invited to choose to converse with me in languages that they were comfortable using. Michelle and Diana chose Chinese to express their ideas during interviews and reflective conversations, while Amber illustrated her flexibility by using both languages during these data-collection procedures. As a result, the participants used primarily Chinese language and occasionally English during the interviews and reflective conversations. Data translation and transcribing took place immediately after each data-collection event.

Data Analysis

In this thesis, an inductive reflexive thematic analysis approach was adopted to generate rich detail in both the Phase 1 and Phase 2 studies. For case studies, data analysis intends to make sense of what has been important to the study and answer the research questions; it is a process involving examining, categorising, testing and recombining data (Creswell, 2014; Yin, 2014). Maxwell (2012) argues that data analysis should not be regarded as only the process of coding, it includes other important forms such as reading and thinking about raw data, analysing narrative structure and contextual information of the data, and displaying data. Hawe (2000) suggests that data analysis should be considered as an activity as it starts in the prefield phases and interweaves with the inquiry. It involves "self-conscious, systematic, organised and instrumental thought" (p. 100). After data collection, data analysis started immediately with data transcription and translation. Then all the data were coded and displayed through NVivo 11, followed by identifying themes and writing up the results. The following subsections provide detailed descriptions regarding each step of data analysis.

Data Transcription

Almost all qualitative studies need transcription, as data-collection methods usually include audiorecordings of interviews, and field notes (Lacey & Luff, 2001). In the current study, I transcribed and/or translated verbatim all the audiorecorded data, derived from interviews, classroom observations and reflective conversations, after each data-collection event. Both original English transcripts, such as Amber's interviews, and transcripts which were translated from Chinese language into English, were emailed to each of the participants to ensure that the translations fully expressed their ideas during interviews. Participants were provided the option to edit the transcripts and to request a copy of the recordings. However, participants were informed that they were no longer able to edit transcripts after the data-collection procedure at the end of 2018. In addition, peer checking was applied to increase the trustworthiness of transcripts. Another doctoral colleague, who is a native Chinese speaker, was invited to check samples from the transcripts to ascertain the reliability of the transcriptions. She signed a confidentiality agreement before checking the transcribed samples. Furthermore, the non-verbal language during data collection was also documented in field notes.

Data Translation

In Phase 1, all data were collected in English. However, as indicated in the previous section, in Phase 2 some data were in Chinese. I translated the Chinese transcripts into English.

Lincoln and González y González (2008) point out that translation issues continuously happen in cross-cultural or cross-language studies. Researchers sometimes feel the stress of delivering data from a native language context into their academic disciplines. Based on this understanding, two decisions were made to increase the trustworthiness of translations during the Phase 2 investigation. Firstly, where participants' statements were primarily in Chinese language with occasional English used during the interviews and reflective conversations, both original and translated transcripts were kept in my computer to reduce the risks of losing or misinterpreting meanings in translations. Where appropriate, extracts from the interviews and reflective conversations are presented in the thesis in both Chinese and English (see Chapter 5 and 6), allowing readers who can speak both Chinese and English to understand participants' original accounts.

The second decision related to my stance during data-analysis processes. As a Chinese-heritage language teacher myself, my Chinese background and experience of teaching the Chinese language helped me understand participants' experience of assessment in both the Chinese and New Zealand contexts. However, as a non-native English speaker, it was challenging to translate Chinese words, especially those carrying cultural meanings, into English. Therefore, I worked closely with my supervisors to ensure translations of interview extracts revealed the meanings of original transcripts.

Data Coding and Analysis

Braun and Clarke (2006) describe thematic analysis as a flexible method to organise and describe data in rich detail. It is "a method for identifying themes and patterns of meaning across a database in relation to a research question" (p. 79). Compared with other approaches (e.g., a grounded theory approach), thematic analysis can be used within different theoretical frameworks as it does not subscribe to any theoretical commitment (Braun & Clarke, 2006). For instance, thematic analysis can adopt strategies similar to those used in grounded theory but without the aims of developing theory. Braun and Clarke (2006) comment that "through its theoretical freedom, thematic analysis provides a flexible and useful research tool, which can potentially provide a rich and detailed, yet complex, account of data" (p. 78). Furthermore, Braun and Clarke (2019) propose and describe reflexive thematic analysis, which is identified as a fully qualitative approach. In reflexive thematic analysis, the coding process is open and iterative, codes can be generated and developed at any point during the whole coding process. The aim of the reflexive thematic approach is to maximise researchers' engagement in "interpreting data through the lens of their own cultural membership and social positionings,

their theoretical assumptions and ideological commitments, as well as their scholarly knowledge" (pp. 848–849).

This study aimed to explore beginning Chinese-heritage language teachers' conceptions and practices of assessment as they began to teach in New Zealand secondary schools. The reflexive thematic analysis approach provides the flexibility to interpret data regarding participants' experience of assessment and develop deepened understandings toward the research questions.

Under the umbrella of reflexive thematic analysis, an inductive method was employed to identify emerging themes. Inductive analysis refers to the process of reading and interpreting raw data in detail to identify concepts, themes, or sometimes a model (Thomas, 2006). It allows researchers to explore raw data without being constrained by a priori expectations or models. In other words, all the emerging themes are data driven. Furthermore, according to Thomas (2006), themes that are closely relevant to research questions are identified and presented during inductive analysis, and thus researchers need to make decisions to choose important data from the set of raw data. In this study, themes or patterns which were most relevant to answer the research questions regarding participants' conceptions and practices were identified and described during the data-analysis procedure. Braun and Clarke (2006) stress that coding data does not happen in an epistemological vacuum; researchers' theoretical and epistemological commitments influence how data is identified and described. The standards that I used to decide the most relevant themes for this study were based on the conceptual framework for this investigation—three aspects of analysing teachers' classroom assessment practices (see Figure 2.12).

The six phases of the reflexive thematic approach suggested by Braun et al. (2019) were adopted to code data in the present study. These are:
1. Familiarisation with data.
2. Generating codes.
3. Constructing themes.
4. Revising themes.
5. Defining and naming themes.
6. Producing the report. (p. 852)

Based on this process, the following subsections introduce the four major steps of the analysis procedure in this study. It is worth noting that the approach of reflexive thematic analysis and

the conceptual framework were adopted to analyse data throughout both phases of investigation. However, the generated code scheme in Phase 1 was not used in deciding codes and themes during data analysis in Phase 2, as the research contexts were different in the two phases. Specifically, Phase 1 focused on seven preservice teachers' assessment experience, whereas Phase 2 aimed to understand three individual beginning teachers' assessment experience when they began to teach in New Zealand secondary schools. Thus, codes and themes generated in Phase 1 may not have been suitable for analysing data in the Phase 2 study.

Step 1: Familiarisation with Raw Data. All the data were read and reread during the data transcription and translation process, and thus some initial ideas were marked or written down. Saldaña (2009) recommends that researchers catch any "codable moments" to precode "significant participant quotes or passages that strike you" (p. 16). During the process of transcribing and translating data, any outstanding features revealing participants' conceptions and practices of assessment were coded; memos regarding key ideas were recorded on the margins of transcripts.

Step 2: Generating Codes. After the process of data transcribing and translation, all the transcripts, audios, memos and documents were uploaded onto a qualitative analysis programme called NVivo 11 for the first cycle of coding. At this stage, interviews and observation data were processed in different ways. For interview data, a line-by-line coding strategy was adopted to generate detailed information about data. Charmaz (2006) indicates that the line-by-line coding forces researchers to keep open to the data and look at the nuances closely. Furthermore, it is also useful to prevent researchers from over interpreting participants' worlds with their preconceived notions so that they can critically think about the research object (Charmaz, 2006). In this study, the line-by-line coding strategy helped to avoid over interpretation of the data. As a result, 48 codes were found in the transcript of the focus group interview in Phase 1, while 98 codes were found in the interview data from different stages of data collections during the Phase 2 investigation (see the example in Appendix J).

In contrast, observational data were coded using a different approach. The line-by-line strategy was not suitable to code the observational data, given that not all the participants' classroom practices during observations were assessment. Therefore, the conceptual framework of this study (see Figure 2.12), with reference to the field notes and memos, was used to identify and code participants' assessment practices in classroom activities. Miles et

al. (2014) indicate that the data condensation process happens when researchers make decisions regarding conceptual framework, research questions, cases, and data-collection approaches. As introduced above (see Figure 2.12), the conceptual framework of this study includes three aspects of recognising and understanding assessment in classrooms. The first aspect, planned and unplanned assessment, was used to recognise the participants' classroom assessment activities. This perspective was built on Cowie and Bell's (1999) framework of planned and interactive formative assessment. Their framework introduces the steps of teachers' practices in both planned and unplanned formative assessment activities. This study expands their framework for formative assessment, categorising assessment activities into planned and unplanned assessment but is extended to also include summative assessment activities. The planned assessment activities include the essential steps of assessment activities, while the unplanned assessment activities include the steps of Cowie and Bell's (1999) model of interactive assessment, including steps of noticing, recognising and responding.

The second aspect, planning, eliciting, interpreting and using assessment data, was used to identify and understand the process that participants use to implement planned classroom assessment activities. However, not all the observed planned assessment activities included these steps. For example, teachers usually plan assessment activities before conducting them in the classroom, and thus the step of planning assessment is not usually observed in classroom assessment activities. Based on this understanding, this study adopted Cowie and Bell (1999) and Brown's (2003) definitions of assessment, in which eliciting, interpreting and using assessment data is considered as the core component of assessment activities. In other words, any classroom activities involving participants in elicitation, interpretation and use of assessment data were identified as assessment activities.

Finally, the third aspect refers to how both summative and formative purposes informed the use of classroom assessment activities. These two purposes of assessment were described in the conceptual framework based on the previous literature (Cowie & Bell, 1999; Harlen, 2007; Rea-Dickins, 2001; Ruiz-Primo & Furtak, 2006). Describing these two purposes of assessment in the conceptual framework aimed to connect interview data and observational data, exploring participants' espoused beliefs and their actual classroom assessment practices in use.

This conceptual framework was used to recognise and code participants' classroom assessment activities during observations (see more detail in Chapter 2, p. 63). After completing the first cycle of coding, initial codes were checked and categorised for generating themes.

Step 3 Constructing Themes. The second cycle of coding aimed to review initial codes and search for emerging themes. Braun and Clarke (2006) conceptualise themes as capturing "something important about the data in relation to the research question and represent[ing] some level of patterned response or meaning within the data set" (p. 82). Braun et al. (2019) further explain that themes are not generated from raw data, the process of constructing themes is to test the relationship between theme candidates, which generate from the first two analytic stages, and the research questions. Miles et al. (2014) identify the second cycle of coding as "pattern coding" (p. 90), where researchers search for themes or categories by summarising initial codes into explanatory or inferential codes.

One of the challenges at this stage was that as data collection proceeded, some of the initial codes needed to be changed or regrouped into different categories. For example, an initial theme called "learn to design assessment" was finally renamed as "experience of designing assessment"; "assessment learning" and "assessment design" being regarded as two subthemes. Braun and Clarke (2013) suggest that as themes may become too diverse or complicated by over interpretation, it is important to identify "what is of interest and why" (p. 92). Terry et al. (2017) suggest that each theme holds important meanings in relation to research questions and thus themes need to be distinguished from each other. They observe:

> Mismatch between what is contained in data extracts, and what the researcher claims those extracts demonstrate (which could reflect poor coding, or poor theme development), either produces a poor-quality analysis, or a headache for the researcher later, when they have to redo the analysis. (p. 27)

Bearing this in mind, the research questions and described conceptual frameworks were used to organise themes in an internally consistent manner throughout the data-analysis procedure. Moreover, documentary data were used as a reference to complement and verify generated codes and themes during analysis. As a result, codes were grouped under 35 themes (see the example themes and codes in the Table 3.5). Codes and themes were reviewed and refined continuously, and a display of the full set data were created to present data systematically (Miles et al., 2014).

Table 3.5

Example of Four Themes with Example Quotations from Stage 1 in Phase 2

Example of themes		Extracts that exemplify themes
Assessment purposes	Using assessments to improve learning	"Assessment is after a period of learning; students could review their learning systematically." (Diana, Interview 1)
		"Students could use assessment as an opportunity to reflect their learning; I could use assessment to know better where they are in learning, and how I should adjust my teaching." (Michelle, Interview 1)
The effectiveness of assessment	Assessment needs to be efficient	"The first one is that the most important things about learning is to be efficient." (Amber, Interview 1)
		"Sometimes when I design activities, I did not expect it could be useless." (Diana, Interview 1)
Prior learning from experience of assessment in China	Benefits of teaching in the Chinese context	"Because in China, teaching English, there is a great teaching and research group. And in China, I came from Beijing, it was completely different, because all assessments are based on the college entrance examination." (Diana, Interview 1)
		"Something like they would plan for teaching together, all the teaching resources will be offered such as they have teaching resources such as they have assessment papers worksheet." (Amber, Interview 1)
Professional development	Rare opportunities for PD as begin teaching	"However, I think this kind of opportunity is still rare. Once or twice a year?" (Diana, Interview 1)
		"I didn't participate in any PD this year, and there is no such kind of PD for Chinese language teachers in our school." (Michelle, interview 1)

Step 4 Defining and Reviewing Themes. At this stage, Terry et al. (2017) suggest that researchers move their position from summarising data into themes to an interpretation orientation. In other words, researchers need to make sense of the categorised data at this stage. They further stress that at this stage researchers need to make sure of the clarity of each theme and its coherence with central ideas of data analysis.

In this study, strategies of *noting themes* and *making contrasts/comparisons*, suggested by Miles et al. (2014), were frequently employed to generate trustworthy ideas and make inferences from categorised data. Noting themes refers to noting recurring themes in which

codes show similarities and differences to each other. As Miles et al. (2014) suggest, themes are found almost by human intuition during the process of data analysis. However, the process of constructing themes requires researchers to do more work on searching for additional evidence of the same theme and keeping open to any emerging codes. Guided by this strategy, for instance, codes such as "NZ system requires assessment capabilities" and "no established resource database in NZ system" emerged at the last stage of data collection. They were finally categorised under the theme of "NZ assessment system," which frequently occurred across the interviews at different stages and was used to explain participants' understandings of the New Zealand assessment system along with their accumulated experienced in teaching in secondary schools (see details later in Chapter 5).

The strategy of making contrasts/comparisons was used for making comparisons within cases and across cases in this study and to increase transferability of generated themes and make inferences accordingly. As Miles et al. (2014) point out, "the *practical* significance (through comparisons of data) is what you need to assess" (p. 249). This strategy was frequently used during the iterative cycles of analysis to compare the conceptions of each participant over time (see example, Table 3.6).

Table 3.6

An Example of Comparing Themes Across the Two Phases Regarding the NZ Examination System in Amber's Case

The example theme	Phase 1 study	Phase 2 study			
		Interview at Stage 1	Interview at Stage 2	Interview at Stage 3	Interview at Stage 4
NZ examination system	Good, small complex system, less stress	Big workload, lacking resources, no ready-to-use resources	The standards of bursary and scholarship (the examination system prior to NCEA) are more useful	N/A	Less guidance from NCEA standards on preparing teaching and assessment

Step 5 Producing the Report. At the final stage, all the themes were assessed with reference to the research questions. Also, analysed data were reviewed to search for any data missing during the data-analysis procedure. Terry et al. (2017) urge researchers to take the last chance to check structures, contents and names of themes, ensuring that themes are close to data and answer research questions. Finally, the reviewed themes are presented and discussed in the following chapters.

The Analysis of Documentary Data

In the current study, both public records, such as class course outlines, school assessment policies, and participants' private records, such as assessment notes, were collected and analysed. Bowen (2009) notes that thematic analysis is also a method to recognise patterns in documentary data. Through carefully reading and rereading documentary data, researchers explore codes and themes in the process of analysing documentary data. During the process of data analysis, public and private records collected at each stage were uploaded to the software NVivo and then analysed by following the five major steps of reflexive thematic analysis (see the example Appendix K). It is worth noting that documentary data were not always coded line by line, information regarding participants' assessment conceptions and practices and factors that may influence their conceptions and practices were coded for triangulating interview and observation data.

Trustworthiness of the Study

While reliability and validity are discussed to ensure the quality of quantitative research, the quality of a study focusing on qualitative methods is demonstrated through its trustworthiness (Flick, 2014; Lincoln & Guba, 1985). Lincoln and Guba (1985) propose four criteria for researchers to examine the quality of qualitative studies: credibility, dependability, transferability, and confirmability.

Credibility

Credibility refers to whether researchers' presentation of the study is accurate and believable (Merriam, 2009; Miles et al., 2014). Edwards (2017) mentions that readers need to know "how well the data and processes of analysis address the intended focus" (p. 133). In the present study, strategies of triangulation, reflexivity, and peer checking are adopted to ensure the quality of the collected data.

Triangulation involves using various methods to verify and generate credible data (Creswell & Poth, 2018; Edwards, 2017; Eyers, 2014). In this study, beginning Chinese-heritage teachers were interviewed and observed at four stages across their first 2 years of teaching. At each stage, interviews were carried out and analysed before observations and reflective conversations were conducted. Therefore, unclear parts in the interview data were collected and participants asked for further clarification during reflective conversations. Furthermore, collected data were also checked during the process of documentary analysis. For instance, participants' statements during interviews regarding tight teaching schedules were checked in

their schools' calendars to ensure the validity of data and further explore possible reasons that contribute to participants' conceptions.

The strategy of reflexivity was adopted in the data-collection and analysis processes to maximise credibility (Eyers, 2014). Simons (2009) argues that researchers need to reflect on the possible influences of their conceptions, values, and biases on the process of data interpretation as well as the presentation of their studies. In this longitudinal study, the friendly relationship between me and participants was helpful for participants to express their feelings freely. However, researchers are at risk of losing their perspective, understanding the issues from participants' perspectives (Corbin & Strauss, 2008; Eyers, 2014). Therefore, I kept reflection notes and diagrams about each participant's story at each stage after collecting and analysing data. In this way, I could refresh my memory and impressions from time to time across the 2 years of the investigation.

Also, it was critical for me as a beginning researcher to discuss categorised codes and themes with supervisors. Lincoln and Guba (1985) and Eyers (2014) recommend beginning researchers to discuss methodological issues with experts in such areas. To enhance the credibility, I presented the emerging codes and themes to my supervisors in our regular monthly meetings, discussing the validity of coded extracts and corresponding themes.

Dependability

Dependability refers to the procedures of data collection and analysis being replicable in other studies (Boodhoo, 2018; Eyers, 2014; Flick, 2014). While the results of qualitative studies may not be replicable, Eyers (2014) suggests that the processes of data collection and analysis are stable in qualitative studies and thus other researchers could follow such research processes. In this study, all the collected data, including interviews, observations, reflective conversations, public and private documents, were recorded and protected for verification purposes. Furthermore, the potential issue of dependability in translation (from Chinese to English) was addressed by member checking as described above (see more explanation in Data Translation). I also checked the transcripts multiple times at different time intervals to increase the "intra-rater" reliability.

Transferability

While the aim of qualitative studies is not for generalisability, "the degree of similarity between the research site and other sites is judged by the reader" (Lodico et al., 2010, p. 173). Lincoln and Guba (1985) and Eyers (2014) suggest that qualitative researchers could maintain

the transferability of their study by providing sufficient descriptions about research contexts so that other researchers could decide if findings in such studies could be transferred to their contexts. In this chapter, descriptions regarding study design, data collection, and analysis are presented in rich details. More descriptive details are provided in the following chapters to allow readers to judge whether findings and conclusions are transferrable.

Confirmability

Confirmability relates to whether the research process can be protected from personal bias (Eyers, 2014; Lincoln & Guba, 1985). Boodhoo (2018) suggests that confirmability can be achieved by employing strategies of reflexivity and frequent checks on the interpretation of data. As stated above, I kept reflective notes to document thoughts and check the objectives of the current study across the 2 years of investigation. Furthermore, I frequently discussed the process of data analysis with supervisors to ensure the collected data were not overinterpreted.

Ethical Considerations

It is necessary to consider possible ethical issues as qualitative inquiry as a research method brings personal lives into the public (Miller et al., 2012). Furthermore, Kvale and Brinkmann (2009) argue that ethical issues are important throughout the entire process, from the first stage of designing studies, such as interviews, to the final stage of reporting. Ethical approval for both phases of this project was sought firstly from, and granted by, the University of Auckland Human Participants' Ethics committee (UAHPEC) before each phase of the investigation was conducted. As noted earlier, participant information sheets (PIS) and consent forms (CF) were sent to participants in Phase 1 (see Appendix C and Appendix D) and in Phase 2 (see Appendix F and Appendix G). Also, PIS and CF were sent to the Dean and the course coordinator of the ITE programme in Phase 1 (see Appendix A and Appendix B) and in Phase 2 (see Appendix L and Appendix M) to seek permission to conduct the investigation. During the investigation, the participants were free to withdraw from this project without providing a reason. They were free to not answer any questions they did not wish to answer, and they could ask for the recording device to be turned off at any time. Participants were offered the chance to amend or comment on their transcripts, as described earlier in this chapter. They could also receive a copy of the transcript, and a final summary of findings, by indicating this on the consent form.

During the data collection, reflective conversations and interviews were conducted in a quiet place at a convenient time for participants. They were audiorecorded with permission from participants. Participants' identities remained confidential throughout the entire study and all names used in transcriptions and the thesis are pseudonyms. All the information, including interview recordings, was kept confidential and secure, only the researcher and her supervisors had access to these documents. As regards classroom observations, all the participants, mentors, and heads of departments were informed at least one week before the classroom observations took place. Prior to classroom observations, the participating teachers explained my presence and sought agreement from students in classes. As recording the participating teachers' practice was the objective of observations, a portable clip-on microphone was used to reduce the chances of collecting students' voices.

Digital data were stored on the researcher's personal, password-protected computer. Hard-copy data (such as transcripts) was stored in a locked cabinet in my office. Data will be stored for 6 years after completion of this study and destroyed after this time. Data collected will only be used for academic and educational purposes, including potential academic publications and conference presentations.

Summary

This chapter provides an overview of the research methodology and detailed descriptions of the study design. To answer research questions regarding beginning Chinese-heritage teachers' conceptions and practice of assessment, a qualitative case-study approach was employed within the interpretive paradigm. Methods of interviews, classroom observations, reflective conversations and documentary data were adopted to collect data regarding participants' assessment conceptions and practice at different stages as they move toward full teaching registration. A reflexive thematic data-analysis approach was employed to organise and make inferences about the data.

In the next chapter, the findings of Phase 1 are presented. Findings from Phase 2 are then presented in Chapters 5 and 6, in which participants' conceptions and practice of assessment are reported respectively.

Chapter 4. Phase 1 Study: Preservice Chinese-Heritage Language Teachers' Conceptions of Assessment

Introduction

In this chapter, findings from the Phase 1 investigation are presented. As stated above, Phase 1 examined preservice Chinese-heritage language teachers' assessment conceptions and confidence to use assessment at the completion of their 1-year ITE programme. A focus group interview and an inductive reflexive thematic analysis approach were adopted to collect and analyse data. Three research questions guided the focus group discussion:

1) What are preservice Chinese-heritage language teachers' conceptions of assessment at the completion of their 1-year initial teacher education programme?

2) What are the factors influencing preservice Chinese-heritage language teachers' conceptions of assessment?

3) How well prepared do they feel to put their assessment understandings and skills of assessment into classroom practices?

At the end of the 2016 academic year, a cohort of 200 preservice teachers completed their studies in a 1-year ITE programme in Auckland, New Zealand, and were ready to seek teaching positions in secondary schools. Seven preservice Chinese-heritage language teachers from this cohort were recruited successfully to the focus group discussion regarding their assessment conceptions during the 1-year ITE programme. This cohort of preservice teachers had strong Chinese-heritage backgrounds, each completing at least 9 years of compulsory education in China. All of them had studied to teach Chinese language and ESOL (English for speakers of other languages) or social studies in New Zealand secondary schools during the 1-year ITE programme. Some participants had experience of teaching Chinese or English language in China or New Zealand prior to this programme (see Table 3.2 for details).

During the focus group interview, participants were encouraged to use English to recall their assessment experience (see more detail in Chapter 3, p. 79) and pseudonyms were employed for all participants.

In Phase 1, although the preservice Chinese-heritage language teachers came with different prior learning and teaching experience, their conceptions of assessment related strongly to their learning experience in the 1-year ITE programme. In this programme, students learned both the theories and practical knowledge of assessment. Also, preservice teachers were

required to undertake 14 weeks of practica, in which they had opportunities to practise their assessment knowledge and reflect their own practices. By the time the focus group interview was conducted in November 2016, this cohort of preservice teachers had finished all the coursework and practica for this ITE programme and were awaiting their graduation ceremony.

During the focus group interview, the preservice Chinese-heritage language teachers reflected upon their assessment conceptions during the ITE programme. Specifically, most of them had shifted from prior negative attitudes about assessment and demonstrated deepened understandings about both summative and formative assessment. They attributed this to the university coursework. Before they studied in the ITE programme, most participants, influenced by their prior learning and/or teaching experience in China, had conceived assessment as tests or examinations and expressed negative attitudes toward assessment. The participants had gained experience in designing and implementing assessment for both formative and summative purposes during the practicum; however, putting what they had learnt at university into practice seemed to be challenging without appropriate support from associate teachers and sufficient resources.

In this chapter, the seven preservice teachers' conceptions about assessment purposes are firstly presented, followed by their assessment experience during practicum. Next, the chapter presents this cohort's report of how they prepared themselves for implementing assessment practice as they began to teach in secondary schools. Finally, conclusions drawn from these findings are presented.

Preservice Chinese-Heritage Language Teachers' Prior Experience with Assessment
In Phase 1, most of the preservice teachers reflected that they equated examinations with assessment before they began the ITE programme in the New Zealand context. Prior to this programme, they reported, their attitudes toward assessment were mostly negative, such as feeling "self-doubting," "anxious" and "nervous" about assessment. Chloe said that she used to consider assessment as being "everything [that] is done on paper." These ideas about assessment were closely related to the assessment system they had experienced in the Chinese context, in which summative assessment was dominant in schools. All the participants agreed that in that context, summative assessment such as exams served the purpose of selection, and thus students were under pressure to be accountable. Ava described her experience of taking various tests:

> I think the typical assessment in China…different kinds of test, and exams, so the test will be taken … by the end of lesson, or by the end-of-unit, mid-term exam, the end-of-term exam, so a lot of small tests and big tests… Because we have pressure to do the exams, we need high scores to go to the university, or go to high school, to very good primary school.

Similarly, Mia described the influences of 高考（Gaokao）, the entrance examination for universities in China, on student understanding of assessment:

> If you pass, you are there, if you are not, you don't have a second chance and have to wait for another year, that is the one most important assessment test in young people's life. The pressure is quite high.

Agreeing with Mia, Chloe reflected that frequent and high-stakes exams had a dominant place in schools in China. Thus, before the ITE programme, they understood assessment as being only for ranking and selection purposes. Chloe commented that their assessment experience could be seen as "learning from exams."

Michelle reflected on one issue caused by frequent exams. She saw the dissonance between students' need for detailed feedback after assessments and the fact that teachers normally do not provide detailed feedback from exams:

> What drives me most is that I actually wish to have more detailed feedback… because that was what actually helped me with my own reflection, and self-correction. But it's not always the case, because we take some many exams, some of exams are just ignored, some of exams we get scores, but we are not getting sufficient feedback, that's what I actually longed for.

In the extract above, Michelle explained that exam results were mainly used for selection or administration rather than feedback on student learning, and thus students could hardly reflect and refine of their own learning by taking exams.

Despite all the participants' conceptions of assessment being influenced by high-stakes examinations, Amber reflected on her experience of formative assessment in the Chinese context as useful to student learning:

> I do remember that sometimes teachers will get us do other things, I really enjoy, which is having class to take turns to do the speech, on a topic of interests and anything, and we just enjoy it so much.

As Amber reflected, formative assessment was useful for students to reflect on their learning and for keeping themselves motivated in learning. However, Amber added that formative

assessment activities did not happen as frequently as tests in China, and formative assessment activities were sometimes carried out in a summative manner:

> As for the informal, as formative assessment, was pretty much like summative ones, it prepared as sometimes like tests.

Amber indicated that she lacked understanding of formative assessment, as a limited amount of formative assessment activities carried out in Chinese classrooms.

Preservice Chinese-Heritage Language Teachers' Conceptions of Assessment

In the interview, all the preservice teachers agreed that assessment purposes include improving teaching and learning, and accountability of teachers, schools and students. The following subsections present detailed information drawn from preservice teachers' insights into each of these different assessment purposes.

Formative Assessment: Improving Teaching and Learning

All the participants corroborated the idea that teachers could use assessment to find the discrepancy between their expectations of students' performances and students' actual performances, and then adjust their teaching practices accordingly. For example, Mia described her understanding of assessment as "an opportunity to fill in the gaps":

> Assessment for me is more like to fill the gaps, find out where the … student is and what kinds of the outcome I want them to achieve, and assessment gives me chance to check and monitor the data then help us to fill up the gap.

Agreeing with Mia, another preservice teacher, Emma, stressed that assessment not only helps teachers and students understand the extent to which students have learned, it also helps to construct the way that teachers and students can "move forward":

> I talk to students, … performance, analyse their strengths and weaknesses, to see the next approach of where we are going and what you need to do to achieve that.

Furthermore, assessment evidence was not only considered useful to adjust teaching practices but also to meet students' emotional needs. Ava, for instance, reflected on one "good" assessment activity that she designed:

> I have an example for that, because I am doing assessment, internal conversation assessment now, with my students. I give them time to prepare, some of them are not confident, they say, "I'm so panic," I say do not worry. Some of them said, "teacher I would like not to go to the front to do the conversation in front of all class, can we do separately or individually with you?" So, I say "okay if you not feeling confident to

go to the front to do the conversation in front of class, you can come to the outside of the classroom, do the assessment separately."

In the above extract, Ava noticed that some students became uncomfortable to do the speaking test in front of students, and thus adjusted her plan and conducted the speaking test with students outside the classroom. As she explained, "we need to cater for their feelings about assessment, to let them conquer their fear, to be more relaxed on their assessment, to have spaces to have them to perform better."

Nevertheless, during the discussion, formative assessment was recognised by all the participants as the major way to improve teaching and learning. Diana indicated that formative assessment usually happened through the student-learning process, and thus it could happen anytime during the learning process. Michelle stressed that formative assessment is learning oriented, and thus feedback on assessment activities is used for learning and teaching. In particular, from their perspective as teachers, Michelle, Chloe, Mia and Diana clarified that teachers use formative assessment to find out where students are at and then decide what the next step of teaching should be. For example, Diana indicated that formative assessment happened in her classroom every day. She kept checking what students had learnt by conducting formative assessment. Thus, at the end of each lesson, she gained understandings of what students had learnt and what she would be teaching for the next lesson. Similarly, Mia said that she assessed students through oral or visual activities or even games. These activities had formative assessment purposes and aimed to help students find where they were at and set up the learning goals. Amber summarised that formative assessment is interactive and usually conducted by teachers in an informal way. Teachers could make judgements about student-learning performance based on their work.

In the focus group interview, all the participants reflected that they gained awareness and deepened understandings of formative assessment through the ITE programme. Emma and Chloe showed interest in trying formative assessment in the classroom; Michelle had gained understanding of designing formative assessment during the ITE programme. She found that unlike the high-stakes summative assessment that she had experienced before, formative assessment was rich in meaning in terms of what it entailed. Teachers could design assessment for formative purposes with different forms and length based on student needs in different contexts, requiring teachers to be creative. Mia confirmed the strength of formative assessment:

I totally believed that assessment is for learning, it's not something that the teacher or school system try to make something quite necessary pressure into students make your life hard, it is just for learning, help them to learn.

With a different emphasis, Ava commented that learning about formative assessment in the ITE programme prompted her to realise that formative assessment evidence should mostly benefit learning rather than teaching. And in agreement with others, Amber raised the challenge of putting knowledge of formative assessment into practice. She indicated that there was potential for teachers to design formative assessment. However, as novice teachers, they lacked experience in designing formative assessment activities to meet student-learning needs.

Summative Assessment: Accountability of Teachers and Students

During the interview, there was some evidence to show that the preservice teachers' conceptions regarding the purposes of teacher and student accountability were not derived from their current experience of assessment in the ITE programme, but from their prior learning and teaching experience in the Chinese context. Specifically, this cohort of preservice teachers finished their 9 years of compulsory education in the Chinese context, and thus their learning experience there closely related to the high-stakes assessment system, Gaokao (college entrance examination in China). Ava, for instance, who, in addition to 9 years of schooling in China, had 8 years of teaching experience there, reflected that students were under pressure to perform well in Gaokao, and teachers were under pressure to assist students to perform well and achieve better scores in exams. As she commented,

I think the typical assessment in China, I think is different types, different kinds of test, and exams….I think it's the same, cause we have pressure to do the exams, we need high score to go to the university, or go to high school to very good primary school, all the kids kind of need to pass the exams, we mainly focus on school.

Chloe shared her experience of learning English language in China. The Gaokao language examination only assessed students in three skills (listening, reading and writing); students' speaking skills were not assessed. In order to achieve in Gaokao, Chloe reflected that when she studied at secondary school, she had to focus on practising language structures and grammar in order to gain better scores in the listening, reading and writing exams. In contrast, as there was no examination in Gaokao to assess speaking skills, Chloe said she did not pay attention to practising her speaking skills. Consequently, she did not speak English frequently, although she was good at grammar.

From the perspective of the teacher, Diana, who had had 11 years of teaching experience in

China, pointed out that she conducted many tests for students in order to provide evidence of her teaching quality:

> We give student tests a lot, but it does not mean we like to give them tests, because everyone is doing the same thing, so if you do not do, you will be the odd one [out].

Due to the pressure from her teacher peers, Diana adopted tests to demonstrate her accountability to the school.

Despite the participants having experienced different pressures to demonstrate their accountability, they all agreed that summative assessment is the most common method to achieve such purposes. Summative assessment was recognised by all participants as tests or exams which happened at the end of learning periods, aiming to evaluate student learning for administration or selection purposes. In line with these purposes, summative assessment was often high stakes, such as "one-off" exams. In the interview, participants' conceptions of summative assessment seemed to be closely related to their responsibility to respond to different audiences, such as parents, teachers and schools. Diana explained that as they needed to provide concrete evidence of their teaching to schools and parents, the results of summative assessment showed their accountability:

> No matter in China, or in New Zealand, actually all students from schools …so they believe in tests, actually, so we teachers, responsible for schools, responsible for students, that is the only thing we can do.

Nevertheless, Ava pointed out that students and teachers were not only under pressure of exams, but also motivated by the pressure of exams:

> So, I think the pressure is good for students, as well as for teachers, as the guideline of teaching how to plan for the whole year. What should I do for the internal? … How can I plan my lesson? I think both students and teachers need the pressure [from exams].

Ava reflected that she used to teach English in a primary school in China; although they had many standardised tests to test student proficiency, she would use results of exams to decide if students were ready to move to the next level of learning and adjust her teaching practice to help students achieve in exams. In other words, teachers and students were prompted to reflect and improve teaching and learning when they were under the pressure of summative assessment.

However, Michelle argued that teachers did not always provide feedback after exams due to the selection purpose of exams. Therefore, it was challenging for students to make use of exam results and refine their learning continuously.

During the ITE programme, the preservice teachers also gained understanding of assessment while they were being assessed as students. Michelle noticed that the ITE programme set up several small assessments rather than a one-time examination. She commented that the frequent low-stakes assessments that they took in the ITE programme were more effective to engage students in learning than a one-off high-stakes examination. As she explained:

> I found [assessment] it is quite different from my university courses back in China, here we have more assignments throughout the year, it's kind like evenly spread out of the year, each weighing 10% something, so it's not high stakes, but it keeps you engaging with the course work …

Learning from the setup of assessment within the teaching programme, Michelle realised that adopting frequent low-stakes summative assessment could reduce the pressure on students from taking one-off high-stakes exams.

Relationship Between Formative and Summative Assessment

During the focus group interview, the preservice Chinese-heritage language teachers also discussed the relationship between summative assessment and formative assessment in the classroom. All the preservice teachers agreed that summative and formative assessment are not opposite in purposes; four participants provided insight regarding the connections between formative and summative assessment. For instance, Diana highlighted the importance of summative assessment, as it decided how formative assessment would be used in the classroom. During the practicum, Diana used the NCEA standards to guide her practices of planning and conducting formative assessment activities in the classroom:

> I think NCEA … determines what kind of assessment we are using in class mostly, because that's what school expect from you, how school checks you, and because that the interest of students.

However, Diana noted that information collected from formative assessment should be used to feed back into summative assessment design:

> I see negative examples in my second practicum, that teacher in the middle of the term, started to prepare for assessment. However, students are not ready for it, actually didn't learn enough to get that done. [Thus] it just sacrificed assessment for assessment,

yeah, maybe she did not know, she just wants things done, and it couldn't be done well.

During the 1-year ITE programme, Diana noticed the dissonance between formative and summative assessment when she undertook her practicum placements in New Zealand secondary schools. She observed that her associate teacher started to prepare students for the upcoming test in the middle of the term, although some contents of the test were things that she had not taught yet. Diana realised that teachers needed to design summative assessment based on the information collected from formative assessment, so that students' learning would not be compromised.

Emma thought that the NCEA standards provided the learning objectives for students to strive for, and formative assessment was used to prepare students for the NCEA exams. For instance, Emma reflected on her own practices of "formative" assessments frequently in order to prepare students for the NCEA exams:

Everyone can achieve something, and everyone can see that "okay maybe this time I can go this far, and then I can answer all the three questions like that," so, they would not feel so intimidated when they actually go to the exams.

Interestingly, her understandings highlighted that she seemed to conceive formative assessment as specific classroom events or activities rather than a formative assessment purpose. In contrast with Emma, Michelle indicated that as formative and summative assessment are not opposite in purpose, they should be integrated together in order to support both teaching and learning.

Summary

In this section, the preservice teachers' conceptions of assessment and their conceptions of assessment have been presented. Firstly, all the preservice teachers agreed that assessment was used to serve both summative and formative purposes. Moreover, in their ITE programme they gained understanding of different types of assessment. Specifically, formative assessment was used for teachers and students to reflect on and adjust learning and teaching. As regards summative assessment, the preservice teachers' responses highlighted the evaluation purpose of student learning. Furthermore, although all the preservice teachers agreed that teachers and students were under the pressure of summative assessment, two preservice teachers argued that the pressure from summative assessment motivated students and teachers to improve learning and teaching practices.

Secondly, the preservice teachers reflected their understandings of the relationship between

formative and summative assessment. Most participants agreed that formative and summative assessment are related. Specifically, some thought this was through feedback, while others thought it was about practising assessments prior to the summative ones.

Influences from the ITE Programme

In Phase 1, although all the participants came with strong Chinese backgrounds and their prior learning and teaching experience from China which influenced their conceptions of assessment, they concurred that the 1-year ITE programme had a great impact on changing their understanding of assessment. In particular, these preservice Chinese language teachers learned about assessment through their university coursework and their two practicum placements in New Zealand secondary schools. In this section, four themes regarding factors influencing these changes in conceptions and practices of assessment are presented: learning from the coursework, accumulated experience of assessment during practicum, influences from associate teachers, and, the importance of resources for teaching and assessment.

Learning from the Coursework

In the focus group interview, the preservice teachers reflected how the ITE programme helped them understand assessment. As stated above, all the preservice teachers agreed that they had deepened their understandings of assessment, formative assessment in particular, during the Teaching Languages course. Through several sessions about assessment in this course, all the participants claimed that they learnt about assessment theories. As Ava commented:

> During my first practicum, I was focusing the planning, classroom management, delivery, I also got chances to explore assessment …Then I learnt a lot from [the lecturer of this] class… so I can see, oh, that make sense now, then I'm going to try more like that [assessment strategies].

Furthermore, the participants reflected that they had learnt assessment through the course structure of the ITE programme. According to the structure of the programme (see Chapter 3, p. 76) the preservice teachers needed to undertake two practica as part of their Teaching Languages course in the programme. Most of the participants found this arrangement was helpful to understand assessment. After the first practicum, participants were provided with opportunities to reflect on their experience of assessment during the practicum and then started to learn theories of assessment. Guided by the lecturer for Teaching Languages, participants attempted to make connections between assessment theories and their practice of assessment and further build up confidence to practise assessment in the second practicum.

Emma, for instance, described her experience of learning about assessment in the course as like "watching a movie":

> I kind have like, when you watch movie, okay, you watch something. And after that you may learned something about that. you can reflect your experience, so it comes more concrete, not abstract, and then I learnt a lot … [in] my second semester, so I can see, oh, that make sense now, then I'm going to try more like that.

Emma found that learning about the theory of assessment after the first practicum helped her make sense of her own practices and led to deeper understandings of assessment theory. Furthermore, she had another opportunity to try what she had learnt in practice when she undertook the second practicum.

However, not all the preservice teachers agreed with the arrangement of "practice before learning theory." Michelle, for example, showed her confusion about this arrangement:

> In practicum one, we didn't learn about assessment, but we do like engagement with assessment… but for me, it was kind of confusing, a little bit, I'm just not sure I'm doing the right thing, and also not that clear of entire purposes or process of assessment.

Michelle wished she had been able to learn about assessment theories before the first practicum, so that she could use what she had learnt in the course to guide her practice in the first practicum.

Nevertheless, although all the preservice teachers agreed that they had deepened their understandings of assessment from the arrangement of interweaving assessment practice with learning theory, some of them still encountered difficulties putting their knowledge of assessment into practice. As Mia commented:

> What I found, probably from my experience, although we have a framework here, the standards website, but practice, resources, the way, how to assess, it all depends on different schools and different teachers assess it, it's all sort of individual decisions, so this means, it is not consistent.

Mia found it difficult to put assessment knowledge into practice, as influences from the practicum school context, associate teachers' conceptions of assessment and experience all impacted on how preservice teachers designed and implemented assessment.

Accumulated Experience of Assessment During Practicum

One question in the focus group was designed to delve into the sources of the seven preservice teachers' conceptions of assessment. Their accumulated experience of assessment during the

two practica was indicated as an important factor influencing their conceptions. Four participants mentioned that they had gained understandings of assessment through their practice of marking tests and giving assessment instructions during practicum. Emma, for example, indicated that she learnt that teachers could gain more understanding of students' learning performances through the practice of marking assessment papers:

> So, when I was marking…. I can see …students' abilities and their performance. Because sometimes their performances did not reflect their abilities, and it helps me know the gap, and explain to the students, let them see "oh, I could do that," it is not "I can't do it."

Marking students' exam papers helped Emma to understand their language proficiencies and learning gaps. Then she was able to provide detailed feedback and build up their confidence for future learning.

Michelle mentioned that she gained understanding regarding assessment through her experience of giving assessment instructions. As she reflected, she only provided brief instructions before tests during practicum. However, students' feedback prompted her to think about the importance of giving more detailed assessment instructions:

> Students' feedback to me, they said "we enjoyed your teaching, and your assessment was good, but if we could have something before the assessment happening, like you are doing the similar questions with us, give us a clear idea what is going to happen, it would be more appreciated."

The feedback from students triggered Michelle to think about effective methods to engage students in assessment. She learnt that providing clear assessment instructions and expectations to students could effectively prepare them for tests.

Influences from Associate Teachers

In the focus group, the preservice teachers shared their perceptions of both the positive and negative influences associate teachers had had on their conceptions of assessment during practicum. From a positive perspective, two of the participants shared that they learnt much about assessment from observing their associate teachers' classroom practice. Ava, for instance, reflected that she learnt how to give assessment instructions from her associate teacher. Compared with sending exam-marking criteria to students and then giving brief instructions about exams, Ava's associate teacher read through the marking criteria and explained what the assessor's expectations would be of students' performances in exams. Ava

realised that giving detailed instructions for assessment is important for teachers to show students what success looks like before they undertake assessment tasks.

Michelle learnt to use formative assessment from her associate teacher. During the first practicum, she observed her associate teacher's assessment practices and found her practice of formative assessment in the classroom was "inspiring." Michelle reflected the suggestions regarding assessment from her associate teacher:

> She said that my classroom teaching is quite engaging, but she found that I need more assessment. She said, "Just bring your mind about the concept of assessment, it's not like something you have to be formal to be summative." She said, "every time you get prior knowledge recycled, just with a little game, or with a little sum up … at the end of the lesson, it could be the assessment of students' learning during this lesson. So, if you have this in mind, keep recycling your knowledge, then the assessment process will be throughout your teaching for your students."

The associate teacher encouraged Michelle to use assessment formatively throughout the teaching process, so that both students and teachers could "recycle prior knowledge."

Although two of the preservice teachers reflected on the positive influences on their learning from associate teachers, another four preservice teachers believed that their practicum experience had not been very positive in helping them understand and practice assessment. These four participants noted that they were assigned to create many resources for teaching and assessment during the practicum, but they had not received enough support to create resources, as Amber commented:

> So much paperwork needs to be done, and there is not enough support, like if they could give us some sort of prescription, not prescription, just suggestion.

Similarly, Mia indicated that although she created many resources for teaching and assessment, and shared with the associate teacher, her associate teacher was reluctant to share her resources regarding assessment with her:

> I feel like [the associate teacher does] not focus on my learning; she uses my time to create new idea for them, so I wouldn't mind sharing actually… I didn't get anything from them…I think in the future, we need to somehow to share.

Mia pointed out that learning from well-established assessment activities could help them learn assessment; unfortunately, she did not receive full support from her associate teacher. Fortunately, as Mia commented, this cohort of preservice teachers in the interview held regular meetings during the two practica, discussing and supporting each other to design and

implement assessment activities. Mia commented that peer discussions and sharing resources was helpful in learning assessment:

> So, I think in the future, we might have support from each other, but I just mean at this moment … you know if you are lucky, you are in a group of lovely classmates, then you will benefit, otherwise you are by yourself…

Overall, the preservice Chinese-heritage language teachers encountered challenges putting what they had learnt in the coursework into practice. Compared with learning assessment from associate teachers, most of preservice teachers received more support with assessment from their peers during the practicum.

The Importance of Resources for Teaching and Assessment

When the preservice teachers were asked to compare their prior experience of assessment and their experience regarding assessment during the practica, three preservice teachers highlighted difficulties in designing assessments. As Diana said: "We still lack resources, I create everything by myself, it took me a lot of time."

Diana found that the difficulty of creating assessments was derived from the fact that she was lacking resources to design appropriate assessments to use in class. Ava mentioned that there were limited open and free resources for preservice teachers. In agreement with these two participants, Mia thought that as Chinese language as a subject was quite new in the New Zealand context, the development of the subject was still underway. As she explained:

> Mandarin is a quite new subject, I mean in New Zealand, I think ESOL or social studies or other subjects got much more online resource, even Japanese and Spanish had more, Chinese hardly any we can sort of use, you can look the idea, but you still need to create yourself.

Mia said that she discussed her thoughts with an experienced Chinese language teacher in a professional development workshop. They agreed that the subject of Chinese language was in progress in the New Zealand context, and thus resources for creating assessment activities for teaching Chinese language were still not enough. Mia suggested that building a database that included experienced Chinese language teachers sharing experience and resources would be helpful for beginning teachers like her to learn and use assessment in the future.

Limited Time Allocation for Chinese as a Subject

During the focus group discussion, Michelle mentioned she only had time to provide feedback on common mistakes in exams, and said that finding how to give feedback to individual students about their future learning was challenging for her:

> But what's tricky to me…is how to cater for the students' needs after the assessment, because sometimes I might only have time to address common area to my class, but how could I help individuals to a further extent.

Ava also pointed out the importance of using summative assessment results to cater for students' learning needs. However, given that Chinese language was an optional subject in schools, there were less frequent classes for students than in other subjects. The challenge for this cohort of preservice/beginning Chinese-heritage language teachers was to use assessment evidence in a tight timeframe of teaching Chinese language in the New Zealand secondary school context.

Summary

In this section, influences that impacted preservice Chinese-heritage language teachers' conceptions of assessment have been outlined. The cohort of preservice Chinese-heritage language teachers found that the setup of the ITE programme which interwove assessment theory learning and practicum experience was helpful for them to understand assessment. Furthermore, during practicum, participants had deepened their understanding about assessment with accumulated experience of using assessment in the classroom. Also, associate teachers were found to have influenced their learning both positively and negatively. Associate teachers' assessment conceptions and practice influenced their guidance of preservice teachers' practices of assessment. Finally, limited resources and less frequent classes for teaching Chinese language were found to have constrained the preservice Chinese language teachers' ability to learn about assessment continuously.

Preservice Chinese-Heritage Language Teachers' Confidence Level for Using Assessment

The responses from most preservice teachers during the focus group highlighted challenges in designing summative assessment and giving students feedback, although some participants showed confidence to implement assessment. Michelle, for example, stated that she became more confident in marking exam papers and Chloe felt confident in observing students' performances during formative assessment activities.

Four participants reflected that designing summative assessments was still challenging for them at the exit of the teaching programme. For instance, Michelle commented about test design:

> It involves a lot of considerations, like what content you are going to put, what format, how many words are they going to write, what is the standard. Who can tell you the standard? I am not sure.

Three factors were found to influence the preservice Chinese-heritage language teachers' confidence in designing summative assessment. Firstly, the participants agreed that they were still lacking experience in designing summative assessments in the New Zealand context. Furthermore, the preservice teachers believed that different school requirements for summative assessment would shape their practice of designing assessments. For example, Chloe and Diana stated that they would adjust their practices of assessment in order to meet the requirements of the school assessment policies.

Finally, Michelle mentioned that she was not confident to interpret collected assessment data:

> What's tricky to me, is how to interpret the information or data I collected from students, how to cater for the students' needs after the assessment.

Michelle reflected that she was not confident in making use of collected assessment information. She connected her interpretation of assessment data with providing effective feedback on student learning, pointing out that teachers need to use assessment evidence to cater to students' learning needs.

In summary, how well the participants prepared themselves in the teaching programme has been examined and presented in this section. Preservice Chinese-heritage language teachers reflected that they were still lacking confidence to design assessments, interpret collected assessment data and use assessment evidence to cater to student-learning needs. Factors such as lacking experience of practices and resources constrained the participants' confidence at using assessment.

Chapter Summary

This chapter has presented the findings in response to three research questions of the Phase 1 study:

1) What are preservice Chinese-heritage language teachers' conceptions of assessment at the completion of their 1-year initial teacher education programme?

2) What are the factors influencing preservice Chinese-heritage language teachers' conceptions of assessment?

3) How well prepared do they feel to put their assessment understandings and skills of assessment into classroom practices?

The aim of the Phase 1 study was to gain insight into preservice teachers' conceptions of assessment as they began to teach in the New Zealand context.

The findings demonstrate that the preservice Chinese language teachers had enriched their understandings of assessment during the 1-year ITE programme and knew that assessment could be used to serve both formative and summative purposes. It was also found that this cohort of preservice teachers had changed their conceptions from their prior understandings in which they understood assessment as summative assessment. However, the evidence from Phase 1 was not sufficient to understand whether these preservice secondary teachers understood formative assessment as a pedagogical process or whether, as is often the case in New Zealand secondary schools, they interpreted formative assessment to mean tests and activities to practise for examinations and high-stakes tests and performances.

In Phase 1, learning about assessment theory, accumulating experience of assessment and learning from associate teachers was found effective in deepening the preservice teachers' conceptions of assessment. However, they believed that a lack of resources and support from associate teachers constrained the preservice teachers' development of assessment understanding and practice. Moreover, their prior experience of assessment in the high-stakes assessment context in China may have still been impacting their conceptions of assessment. Finally, most of the preservice teachers were not confident about designing assessments and giving feedback. Factors such as lacking assessment experience in specific schools and the limited timeframe for the subject of Chinese language subject were outlined.

Chapter 5 Phase 2: Beginning Chinese-Heritage Language Teachers' Conceptions of Assessment

Introduction

In this chapter, research findings from Phase 2 are presented. The conceptions of assessment of the three beginning Chinese-heritage language teachers were examined as they began teaching and moved toward full teacher registration. After the focus group interview in Phase 1, which was conducted at the end of the 2016 academic year, the cohort of preservice Chinese-heritage language teachers graduated and launched their teaching careers in New Zealand secondary schools. While they were seeking teaching positions in secondary schools, they were also invited to participate in the Phase 2 study. Three teachers, Amber, Diana and Michelle agreed to continue to participate in Phase 2 as they began to teach (see demographic details in Table 3.2).

In the New Zealand context, beginning teachers are required to complete a broad-based programme of induction and mentoring while provisionally registered (normally for 2 years) to gain a full practising certificate (Teaching Council of Aotearoa New Zealand, n.d.). Phase 2 explored how the three beginning teachers conceived of assessment and implemented assessment practices in each specific secondary school context during their first 2 years as provisionally registered teachers.

In order to answer the research questions regarding beginning Chinese-heritage language teachers' conceptions and practices of assessment during the 2 years of provisional registration, Phase 2 consisted of four data-collection stages in which participants' conceptions and practice of assessment were investigated through semi-structured interviews, classroom observations and reflective conversations (see Figure 3.2 for the procedure of data collection). Towards the end of the study, in August 2018, one of the beginning teachers, Diana, left teaching for family reasons and thus the final two observations and interviews with her were not conducted.

This chapter is organised around aspects of the participants' conceptions of assessment at four stages during their 2 years of provisional registration. Specifically, these aspects were their conceptions regarding assessment purposes, formative and summative assessment practice, and influences that contributed to their conceptions of assessment. Finally, a brief summary is presented.

Stage 1: Assessment Conceptions at the End of the First Year of Teaching (December 2017)

This section presents the three beginning Chinese-heritage language teachers' conceptions regarding assessment at Stage 1 in Phase 2. Stage 1 was the first year in which the three teachers taught in New Zealand secondary schools. After graduating from the ITE programme, Amber, Michelle and Diana found language-teaching positions in different New Zealand secondary schools and began teaching in February 2017. Amber and Diana taught both Chinese language and ESOL, while Michelle taught Chinese language, Chinese literature and social studies. At the end of the academic year, December 2017, these three participants were invited individually to the first interviews, sharing their assessment experience during the first year of teaching in New Zealand secondary schools.

At Stage 1, the three beginning Chinese-heritage language teachers' assessment conceptions highlighted their understandings of different assessment purposes as well as their conceptualisation of formative and summative assessments as they began to teach in secondary schools.

Purposes of Assessment

During the interviews at Stage 1, the teachers were asked about their understandings of the purposes of assessment. Three purposes—improvement, accountability of teachers and students, and classroom management—were gleaned from their responses. In general, responses from Diana and Michelle in this first interview highlighted their conceptions of using assessment for improvement and accountability purposes, while Amber's conceptions of assessment focused more upon the purpose of using assessment for classroom management.

Improvement. Two subthemes of *improving student learning* and *improving teaching and assessment* were found under the theme of *improvement*. Firstly, the three participants agreed that assessment was valuable for improving student learning. They agreed that assessment was effective in motivating students' learning and building up their achievement and confidence in learning. For instance, Amber found that assessment methods such as interactions in Chinese language were helpful to cultivate students' interests in learning Chinese language. Diana employed the method of translation to assess students' proficiency in vocabulary based on the positive feedback from students. Michelle actively tried different methods of classroom assessment in order to engage students in classroom learning.

Furthermore, while both Diana and Michelle indicated that assessment provided opportunities for students to review their learning and make further progress, Amber did not mention this purpose. Michelle, for instance, provided an example of how students could learn from the assessment:

> 比如说我在 9 年级汉语每节课开头的 5 分钟，大多数课的开头的前五分钟，我们会做 brain storming，把你们到目前为止学到的词汇，想一些，最有关的出来，写汉字，写到自己的本子上来，就是像一个发散的 dictation 一样，然后这个我也并不会收上来，我会在下面走来走去的看，或者他们有问题的时候我也会回答，然后他们就会自己回去去翻自己的笔记本，看看自己写的对不对，然后这些也是很小的 assessment，是给学生自己一个回顾的机会。

> Students are asked to do brainstorm in first 5 minutes in the Year 9 class. They will do brain storming. They need to write down every interrelated word they have learnt so far, it's like a divergent dictation. Their writing will not be collected, what I do is walking around, answering questions. So, they could have a chance to review their writing when they are back home. This is a small assessment, an opportunity for students to reflect on learning. (All translations from Mandarin are by the author)

Michelle illustrated that every assessment activity in the classroom had its purpose of evaluation. Students could grab these opportunities to review how well they had learnt and improve their own learning accordingly.

Similarly, Diana described a successful assessment activity of translation, in which the student made progress in speaking Chinese by making mistakes in this assessment activity:

> 我说一个英语句子，他就以他的英语思维翻译出来，然后我直接告诉他错在哪里，然后他马上就知道了原来这是汉语和英语的不同，然后就会在他的错误上使劲的提醒他一下，不过这种提醒两次三次之后，他的汉语语序就会对了。

> I said an English sentence, he translated it into Chinese with his way of English thinking. And then I told him where was wrong, and then he immediately learnt that this is the difference between Chinese and English. After reminding him two or three times about the same mistakes, he could use Chinese sentences in the right order.

Diana pointed out that students needed assessment to gain understandings of what they had learnt, and then they could make changes in language learning.

Secondly, Diana and Michelle reflected that assessment was also useful for teachers to gain understandings of student learning and adjust their teaching practice accordingly. Michelle,

for example, indicated that each assessment activity was an opportunity for teachers to improve teaching quality. She provided an example of adjusting an assessment activity based on feedback from students:

> 还有一个就是如果一个语言点，我觉得他们应该会，但是好几个人就是模糊不清...但是这个活动的目的没有达到。我是不是要改这个活动，并且在不久的接下来，做一个补充的活动。

> The example is that something I thought I have taught them, and they should be good at it, but many of them are still having questions...So, the purposes of designed activities was not realised. I may think of amending designed activities and do something to make up in a near future.

Michelle pointed out that she adjusted the assessment activity because she noticed that students' performance during the assessment activity did not meet her expectations, and thus she would amend the assessment activity and add teaching into the next classes. Thus, all three teachers had an understanding of how assessment is used to assist students to improve their learning in Chinese language. Two also believed that classroom assessment was necessary for improving their teaching. Within the evidence presented here, it seems that these teachers recognised that planned classroom activities were useful for eliciting information about students' learning, and Diana also indicated that within the classroom dialogue, she noticed, recognised and responded to students to help them see their mistakes and make improvements. In addition to these improvement purposes, these teachers spoke of the accountability purposes of assessment, presented next.

Accountability of Teachers and Students. At Stage 1, two of the beginning Chinese-heritage language teachers, Diana and Michelle, mentioned that one purpose of the assessment was for teachers to establish accountability in schools. For example, Michelle indicated that she carefully planned tests and classroom assessment activities to align with the school assessment standards and the department requirements in the first year, because she was under the pressure of retaining her position:

> 但是都会有 retention 的压力，就是如果一个语言，他的学生一年比一年少，就可能会影响到这个老师的工作，employment,乃至整个 subject 的生存。

> But we are all under the pressure of retention, that is, if fewer and fewer students will continue learning a language, language teachers' employment, and the existence of the subject may be in jeopardy.

Michelle pointed out that her pressure was derived from needing to still have a job next year, which related to how many students would continue to choose Chinese as a subject the next year.

Similarly, Diana demonstrated her preference for using summative assessments:

> 当然是 summative assessment,因为这是对学校和家长的一种责任，这是我要做的，就是教学的重大任务之一。（笑）
>
> Of course, summative assessment. As it represents a responsibility to parents and school, that's what I have been doing, and it is a big teaching task for me [laughs].

She further explained that not only were teachers under accountability pressure, students were also under stress to meet the expectations of others:

> 这个学校的学生来说，她的父母期望很高，然后他们想在每次考试中都有好成绩，每个学生其实都是照着这个目标去的。
>
> So, for students in this school, as their parents have high expectations, they want to have good grades in every exam, each student is actually learning to achieve this goal.

Diana reflected that she found that students in senior levels (Years 11–13) in her school were very much under the stress of the NCEA examinations. She said that even Year 10 students were under stress as they needed to decide on subjects for the next 3 years. Diana indicated that once students decided on those subjects, they would be motivated to improve their learning performances under the pressure of the NCEA examinations.

Clearly, all three teachers were aware of the accountability purpose and role of qualifications. They explained that, as teachers, they felt significant pressure for students to perform well in tests and examinations. Diana's comments reveal how younger students within a school environment can feel examination pressure, even before they themselves need to take the examinations. And Michelle's remarks demonstrate that teachers are motivated by the need for students to do well, even to the extent of keeping their job.

Classroom Management. In contrast to Michelle and Diana, who believed that assessment could help students make progress in learning and cause teachers to demonstrate their accountability, Amber claimed that she used assessment as a tool to maintain classroom order. In the first year of teaching, Amber described her teaching experience as "chaotic." She thought that she had not had enough time to design and implement formative assessment in the classroom, because the issue of classroom management was overwhelming for her:

My classroom management is also very bad. When I do the classroom assessment, my first consideration wouldn't be learning, it would be classroom management and how do I make these people behave themselves.

Amber reflected that she implemented many activities in every lesson in order to keep students engaged in learning. For example, she designed a picture book, in which she included language features that students had learnt before. Students could use this book to practise language features that they had learned. However, Amber said the picture book was mainly used to keep students quiet and behaving themselves in class.

Interestingly, although Amber claimed that she had no time to do formative assessment activities in the classroom, given that she spent lots of time maintaining classroom order, she was surprised when her mentor gave her positive feedback after observing her classes:

My mentor observed me four times…it was pretty successful…and she has got really detailed report written, I can't remember what I did, but I must have done something really well …but I forget the details.

In the extract above, Amber connected her success with her strategies to engage students in classroom learning, but she could not clarify what element of her practice made her successful.

Beginning Teachers' Conceptions of Formative and Summative Assessments
At Stage 1, the beginning teachers reflected upon and conceptualised their formative and summative assessments. The following subsections provide in-depth details of their conceptions regarding these two aspects of assessment.

Formative Assessment. In the interviews, Michelle and Diana shared similar understandings of the formative purposes of assessment and stated that their formative practices were accordingly guided by these understandings. In contrast, Amber's conceptions and practices of formative assessment changed distinctively at this stage.

Diana and Michelle agreed that formative assessment provided opportunities for teachers and students to continually refine their learning and teaching. For example, Michelle reflected that she usually arranged for students to do a "warm-up" task at the beginning of each lesson, as it was useful for students to review what they had learnt before. Similarly, Diana indicated that formative assessment activities were effective in motivating low-achieving students to continue learning:

然后我发现平常汉语还挺差的，一点 confident 都没有的学生，玩儿起 game 来还挺高兴的...

然后 participation 特别好，积极踊跃，每个人都怕自己落下，我在想，当然有 assessment

在那里，这个不可能完全实现，所以调动那些不是成绩很好的学生的积极性的话，时不时的这个 formative assessment 大概也要来一下。

I found those low-achieving students who are not confident at all are quite happy when playing games... Then the participation was particularly good, everyone is afraid of falling behind. I think, of course there is assessment. So formative assessment is quite effective on motivating low achievers, therefore I will use it from time to time.

Guided by their conceptions of formative assessment, Diana and Michelle tried various activities to cultivate students' interest in learning Chinese language. Diana said that she sometimes asked opinions from students with low achievement when designing formative assessment activities so that students could "more enjoy" such activities. Michelle developed a system of formative assessment activities, at both difficult and easy levels, so that she could adopt assessment activities in the classroom based on student-learning capability. Interestingly, both of these two participants seemed to see formative assessment here in a "planned as an activity" sense.

However, both Michelle and Diana agreed that they were challenged by limited resources and support when designing formative assessment activities. Diana reflected that she used to work on weekends to create formative assessment activities:

有时候你想半天，其实完全不如借来那么容易，借来了试一下行就可以，但是自己的创作能力有限嘛（笑）...所以想活动也很费神，但是我觉得这个还是有点闭门造车的感觉，真的是应该校际之间多交流。

Sometimes designing assessment is very time consuming, it is not as easy as borrowing assessment from other places... After all, everyone has limited creativity. [Laughs]… But thinking up activities is quite like "working behind closed doors." I think there should be more inter-school communications, as I need more support to create activities.

Diana pointed out that there were limited resources for Chinese language teachers and limited communications between experienced and beginning Chinese language teachers. Similarly, but from a different perspective, Michelle found that she needed to build up her database of formative assessment activities, as she did not receive enough support from her mentor. Her mentor, who was an experienced Chinese language teacher, seemed to have little communication with her about existing resources for assessment activities in her school:

有的时候觉得没有得到特别充足的支持...觉得自己缺少一些资源，或者说，没有 full access 一些资源。

Sometimes I feel that I don't get enough support... I feel that I lack some resources, or that I don't have full access to some resources.

Michelle demonstrated that she was confident in designing formative assessment activities; nevertheless, creating formative assessment activities increased her workload, and thus she had to work in weekends in order to build up a system of formative assessment activities, with little help from her mentor.

In contrast to Diana and Michelle, whose conceptions were similar to their ideas in the preservice stage, Amber's conceptions of formative assessment changed distinctively at Stage 1. In the first semester of Stage 1, Amber characterised formative assessment as a teacher-oriented interaction between teachers and students in the classroom:

I know there is also an assessment to point students in class and let him talk. Right, [then] you insert your feedbacks.

In order to maintain classroom order, Amber claimed that she barely used teaching time to assess students. Moreover, Amber pointed out that during interactions with her students, she only picked students who raised up their hands to answer questions in class:

So, I seldom use the teaching time to assess, I can only assess those who raised hands. For those who don't raise their hands, I can't assess them, right? If you assess him, he will mess up.

Amber explained that she chose students who were confident in answering questions, as she hoped that students could provide correct answers so that the classroom order would not be disturbed by group discussions. With time saved from classroom management, Amber said that she could deliver more teaching content in order to ensure the effectiveness of her teaching practices.

However, Amber's plan of teaching intensively in classes did not work well. She found students were struggling to engage in classroom learning and complete assigned homework:

At the beginning, I don't think that's my problem. If I tell you to listen, you don't, you won't be able to use it...then I found it is unrealistic for them to go back practising listening, as they don't have such strong motivations. I can't expect them to learn language like I would, so I try to leave some time for them to do in class, hopefully, once they enjoy their achievements, they can slowly develop their capabilities.

Amber found that the method of providing a great deal of teaching content in each lesson was not effective to help students' learning. Therefore, in Semester 2, Amber had changed her assessment practice in the classroom. She tried to circulate in the classroom and monitor students' learning while they were working on assigned tasks.

At the end of Semester 2, Amber reflected that she had learnt through her practice that assessment was important, and that the purpose of formative assessment activities was to understand student learning and monitor their learning. However, Amber pointed out that her struggles in teaching and assessment in the first year were an inevitable step to realise effective teaching, which was a teaching philosophy that she had brought with her from the Chinese-heritage culture:

> I may make many detours, such as stuffing too much content… but I don't think what I pursued is wrong, I think it's a balanced thing, it may lead to something bad, but still, I think that's my background, that's get from me, is efficiency.

Interestingly, through her struggles, Amber appeared to regard formative assessment as a process of understanding her students' learning and needs. In comparison, it appears that Michelle and Diana had an understanding based more on planned assessment activities which they interpreted as formative assessments.

Summative assessment. At the end of their first year of teaching, the beginning teachers' examples of summative assessment suggested their conceptions related to practical considerations regarding summative assessment, and difficulties in designing summative assessment.

At this stage, the three beginning teachers highlighted their emphasis on summative assessment. Michelle indicated that she put more thought into designing and marking exams rather than formative assessment activities, although she believed that both formative and summative assessment were important. Michelle explained that she was under pressure to design exams based on the school's standards, as designed test papers would be checked by the head of department and seen by parents. Michelle described her experience of designing summative assessment as "带着镣铐跳舞" （dancing with shackles).

Similarly, Diana pointed out that her assessment practice was influenced by the school assessment environment and her capability to design effective assessment activities. First, Diana indicated that she had changed her assessment practice, as summative assessment was dominant in her school:

平常的时候，肯定会有考试的负面作用，但是对于私校来说，我感觉，家长还是很支持的，他们的初衷还是让孩子在 academic 上有更大的成就吧。所以我感觉还是和整个学校的性质，和学生家长的要求是有关的。

> Normally, the negative influence from exams was outstanding, but for the private school, I feel that their [students'] parents are very supportive, as their intention is to let the children have greater academic achievements. Therefore, I feel that implementing summative assessment is related to the nature of the school and demands of parents.

Diana worked in a private school, in which summative assessment seemed to be very important, as parents and students had the clear objective of achieving in exams. Diana thought the nature of private schools was to assist students to realise academic achievements.

Secondly, Diana actively tried what she had learnt about assessment from the ITE programme in class, but she found that she was still not capable of using assessments effectively:

> 这个还是和自己能力有关（笑），所我感觉，学的是很好的，我也在尽量用，其实刚开始的时候，但是用的用的就走到了学校设计的那条路上了。

> This is also related to my capability [laughs]. I feel that what I have learnt is useful, in fact, I have been trying to make full use of it since the beginning. But I soon found myself back on the road of school-designed assessment.

Diana explained that she did not see school assessments as the opposite of the assessment theory that she had learnt in the ITE programme. In fact, Diana found that her learning from the ITE programme prompted her to reflect on her practice frequently, and she felt more confident in teaching under the guidance of learnt theories. However, within the tight teaching schedule, it was challenging for her to implement assessment theories, such as task-based language teaching and assessment, into classroom practices.

Furthermore, Diana's experience of summative assessment at this stage was influenced by her prior teaching experience under the Chinese assessment system. In the Chinese context, the teaching and research groups in each district of her city were responsible for designing examination papers and organising regular workshops for teacher development. This was both challenging and beneficial for teachers. On the one hand, the examination papers in each district were unified and thus teachers were under pressure to compare examination results with other schools. On the other hand, teachers were not required to be capable of designing

effective summative assessment, as all the test papers were designed by the teaching and research groups. In addition, the teaching and research groups held workshops every week, and thus teachers could develop teaching skills and communicate with counterparts in other schools through observation of classes and panel discussions.

However, when Diana started teaching in New Zealand, she was not confident in using summative assessment, as schools and teachers within them were responsible for designing their own exam papers in the New Zealand context:

> 但是在新西兰，一个学校出自己的考卷。我觉得教师负担多了，就是你要出考卷，你各种能力都要有。

> But in New Zealand, a school is responsible for their own papers, I think teachers therefore have more workload and are required [to have] all kinds of capabilities.

Furthermore, there was only one other teacher, her mentor, responsible to develop curriculum and assessment of Chinese language in her school, so Diana had fewer opportunities than she had had in China to communicate with other in-service Chinese language teachers:

> 教研组很小，只有一个老师，当然你要同西班牙语老师交流。

> The teaching research team is quite small, only one teacher, you certainly need to communicate with the Spanish teacher.

Diana reflected that there were only two Chinese language teachers in her school sector, and that she and her mentor worked well together. However, she wished that she could have more opportunities to communicate and work with Chinese language teachers in other schools.

From a different perspective, Amber's understanding of summative assessment was closely related to the fact that she had less teaching experience than the other two participants. Amber indicated that she was more confident in using summative assessment such as unit tests or examinations than she was with implementing formative assessment activities. This was because, as a beginning teacher, Amber found herself struggling to grasp the flow and pace of teaching and assessment in the classroom. She was always worried about how to have students behave themselves during formative assessment activities, and thus these activities did not work well in her classroom. In contrast, the examination database for Chinese language was well-established in her school, and thus Amber felt more confident in adjusting exam papers from previous examinations and implementing exams in her classes. As she explained, "That's what we are working for. Then it [exams] is better for me to manage."

Amber showed her confidence in adapting previous test papers in order to assess student learning, whereas Diana and Michelle expressed their struggles in designing summative assessments as beginning teachers. Michelle showed her concerns about designing valid test papers under the pressure of job retention and keeping students in the subject of Chinese:

> 但是我觉得在这些方面的资源并不够多，然后我今年从学校提供的角度和我主动是探索的角度没有花那么多的时间精力去参加 PD。

> But still, I don't have many resources to design summative assessment, and I spend my time on exploring how to design assessment rather than participate in PD [professional development] … so my major difficulty is to design summative assessment.

Michelle was hindered in designing test papers effectively, given that she lacked experience of designing test papers and there were few opportunities for her to communicate with her mentor regarding test papers.

As regards Diana, she had trouble in arranging exams based on school calendars:

> 我的 mentor 提醒我，就是咱们要考试了啊，把你的考卷给我看一眼啊。然后我就开始出考卷，然后我就发现有些东西需要补，就比如说有些东西我没有说到的，我后来才知道，我可以根据自己的教学内容改考卷，但是我刚开始的时候也没有这个胆量，为什么呢，因为我怕我教的东西太少，教学目标达不到。

> I did not realise that the test was coming until that my mentor asked [for] my test paper. Then I found something that I have not taught in the classroom during this process of generating a test paper. I later learned that I can change the test according to my teaching material, but I did not have the courage at the beginning. Why? Because I'm afraid I taught too little to achieve the teaching goal.

Diana reflected that as a new teacher in her school, she worked hard on learning the structure of school assessment and tried to complete the required teaching tasks. However, as a new teacher in her school, it was still challenging for her to adjust previous test papers based on her teaching content.

Summary

At Stage 1, all the beginning teachers agreed on the assessment purposes of improvement and accountability, while Amber's conceptions of assessment also highlighted the purposes of using classroom assessment to maintain classroom order. Furthermore, the reflections of the

beginning teachers in terms of practices of formative and summative assessment at the end of Stage 1 indicated their purposes of using assessment included motivating students' learning and that they had an emphasis on summative assessment.

The challenges for the beginning teachers to use different types of assessment were also outlined. Amber struggled with the issue of classroom management, and thus she used formative assessment activities as a tool to maintain classroom order. In contrast, Diana and Michelle struggled to create effective assessment tasks with their limited experience and resources.

It was noteworthy that compared with the other two beginning teachers, who had previous teaching experience, Amber's experience regarding assessment over the first stage changed distinctively. This might be related to the fact that she had had no prior teaching experience, while Diana and Michelle had teaching experience of 11 years and 1 year respectively. Furthermore, although the three beginning teachers spoke of using similar practices of assessment, their reflections at this first stage were mostly about their experience of designing different types of assessment activities and less about how they implemented and used assessment.

Stage 2: Student-Centred Conceptions of Assessment (Term 1 of the 2018 Academic Year)

Data from Stage 2 was collected during the first term of the beginning teachers' second year of teaching. After the first year of teaching, all the teachers reflected that they had accumulated experience regarding both formative and summative assessment and were keen to make changes in their practices accordingly. Thus, the Stage 2 investigation inquired into the ways that the three beginning teachers had made changes to their assessment practice and their corresponding conceptions. Furthermore, as all the beginning teachers had provided little information regarding their implementation and use of assessment at Stage 1, their practices of assessment were explored in more depth at Stage 2. The findings at Stage 2 highlighted the meaningful attempts by the beginning teachers to change their roles in classroom assessment, including encouraging students to take responsibility for their learning and providing effective feedback.

The Relationship Between Assessment, Examination and Testing

In the interviews at Stage 2, all the participants were firstly asked to clarify their understandings of the terminology of *assessment*, *test* and *examination*. This was because

during the interviews at the first stage in Phase 2, they had been using the word "assessment" to refer to "testing" or "examinations" in English sometimes, which might relate to their conceptions of assessment.

The responses of the beginning teachers at Stage 2 highlighted that they had deepened their understanding of assessment from practice. First, all three beginning teachers agreed that assessment was a broad concept, and that exams and tests were two forms of assessment. For example, Diana indicated that assessment was important to moderate the discrepancy between teaching plans and actual teaching practices:

> 比如说老师设计课程是一件很主观的事情，它是建立在我对学生的了解，还有我想让他们学的基础上，但是实际教学效果和自己的想象有的时候是有出入的，有各方面的影响因素。
>
> Designing course is a very subjective thing. It is based on my understanding of the students and what I want them to learn. However, there are discrepancies between the actual teaching effect and your own imagination, and there are various influential factors.

Diana provided an example of an unsuccessful lesson at the end of the last academic year. She said that she overfilled the teaching content in that lesson, therefore students were often off task during classroom assessment activities. After that class, Diana reflected that students' performances on that lesson prompted her to plan to teach less new content and add more activities, so that students could be more engaged in classroom learning.

Secondly, the three beginning teachers held different understandings from each other about tests. Diana indicated that examinations and tests were two forms of summative assessment, and Amber pointed out that a unit test was an informal assessment vis-à-vis formal examinations with large weight and consequences.

Different from the other two teachers' conceptions, Michelle demonstrated her understanding of tests as a form of assessment which integrates the characteristics of both formative and summative assessment:

> 然后我每一个大的 unit 有三个 topic，每一个 unit 教完之后，我那个班上的学生都知道会有一个这样的 test，他们知道这个会作为我对于他们平时表现的成绩的考虑，这个不会是登记到 report 上的成绩，但是他会记录在我自己的 grade book，或者成为我的一个依据，就如果学校有年终评奖或者其他...所以他们一方面很重视，另一方面不会很 stress。

We have three topics in each unit. After learning topics in each unit, students know that there will be such a test. They know that the results will be considered as the results of their participation, rather than the results which will go on the reports. but I will record the test results in my own grade book as a reference for me when the school grants awards for students with good results…so, they pay attention to it on the one hand, and are not very stressed on the other hand.

Michelle explained that on the one hand, students had an understanding that they were assessed systemically after learning a unit, and unit tests were a bit like summative assessment from this perspective. Michelle stressed that students had been informed that the purpose of unit tests was to provide an opportunity to review their own learning rather than demonstrating their accountability, and their performances in unit tests would not go into their end-of-year report. She further clarified her understanding, seeing unit tests as a type of formative assessment:

那怕是 unit test 上想成 summative assessment,从我的角度也会觉得，好像他们得到了一个成绩，然后就到此为止了，我会倾向于，总给他们点小的，嗯，反馈或者是纠正，针对于这些学习的建议，但是我觉得这种 feedback 和 feedforward *存在了*，这就是一种 formative assessment。

Even if the unit test is a type of summative assessment, from my point of view, it seems that they get a score, and then it ends. I will tend to always give them a little, um, feedback or correction for learning, but I think since such feedback and feedforward exist, it [unit test] is a kind of formative assessment.

However, Michelle pointed out that she would use results of unit tests as evidence of students' performances when the school granted awards to students with better performances in unit tests and examinations. In other word, Michelle conceptualised unit tests as assessment for both formative and summative assessment purposes.

Assessment Purposes

At Stage 2, all the beginning teachers' conceptions of assessment purposes were in line with their understandings at Stage 1, illustrating that assessment was used to fulfil the purposes of improvement and accountability. However, the collected data at this stage highlighted two purposes, adjusting teaching and assessment, and demonstrating student accountability in their conceptions. The following subsections provide detailed descriptions regarding the beginning teachers' understandings of these two purposes of assessment.

Adjusting Practice of Teaching and Assessment. In the interviews, when the beginning teachers were asked the purposes of assessment, their responses indicated that they frequently used assessment to adjust teaching or assessment. For Michelle, teaching plans could serve the learning needs of most but not all students, and so she used assessment to understand individual student's learning needs and then provide learning suggestions accordingly. Diana used assessment to inquire into the extent to which students had learnt in class and then adjust teaching content accordingly. In this way, students could continue engaging in learning in the classroom. Furthermore, Diana also adjusted assessment methods based on the responses of students. She reflected that she used to take Year 7 students to computer labs for practising language features on computers, as she found students in Year 7 needed more "games" to motivate them to continue learning the Chinese language. However, the method was not effective for Year 10 students, as they had enough motivation to learn Chinese language.

As regards Amber, she had realised that assessment was an important element in her teaching practice at the end of first year. In this interview, when she was asked about the purpose of using assessment, Amber reflected on an example of conducting a reading task in class:

> 还有我布置得 reading task，有几个 questions 他们要做，然后我 observe 他们怎么样，有没有 challenge，是不是 too challenging... 所以我 observe 的第一目的就是为了知道这个东西对不对他们的水平，是不是对他们有效，因为如果太难的话，他们就会彻底 switch off，那样我就需要提供其他形式的帮助，这是第一个目的。嗯。。第二个目的就是说，嗯。。也想知道他们那些会做哪些不会，比如说他们 9 题做了 8 题，那我可能就不用花时间讲了，对不对，文章如果他们只对了一两题，我可能花点时间讲讲里面的句子.

> I have a reading task, there are a few questions they have to do, then I am observing them, thinking whether there is a challenge, or if it is not too challenging … so the first purpose of my observation is to know if this thing is right for them, is it effective for them, because if it is too difficult, they will completely switch off, then I need to provide other forms of help, this is the first purpose. Ok. The second purpose is to say, um ... I also want to know what they can do or cannot do, for example, they have answered 8 in 9 questions, then I may not have to spend time talking through these questions, right? If they only have one or two questions, I may take some time to talk. Talk about the sentences in the questions.

By observing students' performances in the reading task, Amber made decisions to adjust the degree of difficulty of the assessment task and analysed the responses of students in each reading question in order to adjust her teaching emphasis.

Demonstrating Student Accountability. All the beginning teachers agreed that one of the purposes for which teachers used assessment was to motivate students to take responsibility for their own learning. For example, Amber reflected that she used weekly vocabulary tests to check student learning and ask them to be responsible for their learning:

> 这个就是给我一个 general idea 他们到底知道哪些词学了哪些词，做了作业没有。我觉得我用 test 主要就是 accountability，如果他们没有好好做作业的话，可能还会有 consequence，比如说 detention 之类的。所以这个 test 对他们来说也是会有一些 pressure。
>
> This [vocabulary test] is to give me a general idea of which words they have learned, and whether they did homework. I think that I use tests mainly for accountability. If they don't do their homework, they may have consequences, such as detention. So, this test will have some pressure for them.

Amber believed that students were more responsible for their learning under the pressure of tests. Michelle agreed with Amber's understanding and pointed out that students would spend more time on preparing for tests than assigned tasks. She was constantly asked by students if results of unit tests would go on their reports, although she had explained that the purpose of unit tests was to give students a chance to review their learning, and the results of unit tests would not be included in academic records.

Diana claimed that students needed to take responsibility during classroom learning processes:

> 就是让他们做一些改变，让他们想办法让他们 take responsibility。我后来因为他们有 writing booklet，然后我就用了一点强势，嗯，你就要在下课之前，嗯，完成到第几页然后那个…也就没有再进行我原来计划的那个部分。练习，然后我 individual check。
>
> Just let them make some changes, let them take responsibility of learning. I later asked them to do writing booklet, um, I was a bit tough, "you have to finish to xxx page before the class is over, there is a writing task in the writing booklet…" Then I didn't go with my original plan. And I have individually checked their homework.

Diana assigned homework to motivate students to take responsibility for their own learning when she noticed that students could not engage in her designed tasks. In this way, all three

participants were using tests and formal tasks, including homework, as ways to motivate their students and encourage them to take responsibility for their learning.

Formative Assessment

At Stage 2, Michelle had established her confidence in creating formative assessment activities, while Diana and Amber were still seeking appropriate methods for conducting formative assessment effectively. Michelle, for example, explained that she had built up a system of formative assessment activities, including both easy activities and comprehensive activities, during her first year of teaching. She was more confident in refining and putting previously designed activities into practice at this stage. Compared with Michelle, Diana and Amber reflected that they had deepened understandings of the effectiveness of teaching and assessment. For Diana, gaining understanding of students' characteristics and learning needs helped her become more sensitive to students' language proficiency so that she could design suitable formative assessment activities. She indicated that she was more confident in preparing each lesson based on students' proficiency of Chinese language. Furthermore, learning from classroom practice, Diana realised that a large amount of language input, in a teacher-controlled assessment environment, was not helpful for student learning:

> 我是想控制完以后再放，在这个 context 下可是我这个 input 不是很成功（笑）。我想 scaffold 但是我这个 scaffold received poorly. ...我先开始想控制，我想更多的 input, 不太适合当下的情况...因为他们已经 capable learner, 可以说出一些东西，如果是由他们自己说出来的，提出的要求，学习语言要比我这样强。

> I was thinking to control first then let them perform freely, building up a context for them to practise, but my input was not very successful [smile]. I wanted to scaffold but my scaffold was received poorly… I want to control first, I want more input, but this was not suitable for the current situation, and I don't need more input. Because they are capable learners and can speak something in Chinese language, they could've had stronger motivations by speaking their learning needs, rather than me [being in control].

Diana explained that teaching in New Zealand was different from her prior experience of teaching in the Chinese context, in which teachers fully planned teaching content and structured assessment activities in each lesson. Diana learnt that large amounts of teaching input and a teacher-controlled assessment environment did not effectively promote student

learning. On the contrary, when they played a central role in formative assessment, students would become more motivated if encouraged to take responsibility for learning.

At this stage, Diana also expanded her social network in the school and learnt many methods of designing formative assessment activities from her colleagues. She said that she and the Spanish language teacher usually exchanged ideas about planning formative assessment activities, and thus she received many practical suggestions about planning assessment. In fact, Diana had expressed her intention of expanding her social network in the school at Stage 1. As she explained, due to her personality, she barely talked to colleagues in the school in the first year. This resulted in her difficulty in seeking advice from others when she had trouble understanding the structure of school assessment. Then she decided to make changes after the deputy of the school suggested to her that "每个人至少要有一个朋友才能在学校立足" (if a person wants to stand in one place, at least have a friend). Diana agreed with her deputy as she found seeking support from colleagues was easier than thinking about teaching solutions by herself.

Amber had been actively exploring effective ways to teach and assess throughout Phase 2. She explained that pursuing effectiveness of teaching was influenced by her Chinese background. Amber used to think teaching effectively meant intensive input and frequent practice of the target language. At Stage 2, Amber shifted her understanding toward effectiveness of learning:

> 如果别人说一个人，想用外语做点什么事？比如说他想，留学，或者是，考雅思，快速学会一门外语的话，找工作的话。但是因为这些学生他们还太小，他们学习的目的和我们那种，学位，一味追求效率的学习方法，不一样，所以我觉得多花一点学习时间也也可以无所谓的，只要让他们明白他们为什么要花这么多时间学习。

> If a person wants to do something with a foreign language, for example, he wants to study abroad, or, to take IELTS tests, or to quickly learn a foreign language, to find a job. [It is the fastest way to do it.] But these students are still too young, the purpose of their study is different from our kind which includes degree and pursuing efficiency. So, I think it is okay to take a little more time to learn. Just let them understand why they spend much time on learning this.

Amber explained that she changed her conceptions as she realised the discrepancy between her teaching goals and the specific teaching context. At Stage 1, Amber reflected that she assumed that young learners did not have as strong self-regulation as adults and thus they needed more support to explore their own way of learning. However, Amber found that

students did not have as strong a motivation to learn the Chinese language as she expected. In her school, most students had their priorities on compulsory subjects, such as maths and English, rather than the optional subject of the Chinese language. Therefore, it was unlikely for her students to achieve the level of Chinese proficiency that she had planned.

After several unsuccessful attempts, Amber adjusted her teaching plans at Stage 2. On the one hand, she tried to cultivate students' self-regulation by sharing effective methods of learning Chinese language. On the other hand, Amber designed many classroom activities for her classes, as she found that these activities seemed to be more effective to engage students in classroom learning:

> 因为那种 format 对他们来说可能太过沉闷了。但是这种做法，他们就学到很多...那不管哪一种，只要有效，能帮助我达到最后的目标就可以。
>
> Because that format may be too dull for them. They did not learn a lot, but they learned a lot from this interesting approach. No matter which one, as long as it is effective and can help me achieve the final goal [of engaging student in learning].

However, Amber was still struggling to create effective formative assessment activities at this stage:

> 我希望有的东西是 classroom ready 的，那种东西是我不知道你以前在国内有没有上学用过那种就是已经装订好的 workbooks(就像有填空回答问题，这样子对吗?) 对对。只可以从里面找一段，你可以布置下来. 但是老师有很大的灵活性，因为他可以选择里面的部分来做。
>
> I hope that there is something [resources] that is classroom ready, such as workbooks when I studied in China. (Interviewer: You mean the workbook with exercises like fill in the blank?) Yes. You can only use a paragraph from inside [the workbook] in the classroom. But the teacher has a lot of flexibility because he can choose the part.

Amber claimed that it was challenging for beginning teachers such as her to build up a data base of formative assessment activities in the New Zealand context. Amber pointed out that beginning teachers in China had more "ready-to-use" resources in order to ensure the quality of their formative assessment practice in the classroom.

At Stage 2, all the participants seemed to conceive of formative assessment as planned activities, given that they explained assessment as planned classroom activities or tasks rather than seeing it as a process such as interactions between teachers and students.

Summative Assessment

At Stage 2, these three beginning Chinese-heritage teachers made meaningful changes in practice, along with displaying deeper understandings of school context and students. Diana, for instance, reflected that she almost "messed up" the first examination in the first year, as she had not completed teaching everything she had planned to when the school examination was coming. At Stage 2, Diana said that she carefully planned examinations based on the school assessment standards. She increased her confidence with accumulated experience of summative assessment:

> 今年比去年更加 well prepared, feel more comfortable and confident about that。我觉得
>
> 经验也是学来的，嗯，看的多了，可能，受挫折多了。
>
> This year I am more well prepared than last year, feel more comfortable and confident about that. I have learnt a lot. Well, as I have seen more. Maybe, as I have been through lots of frustrations.

Amber planned to employ an approach of summative assessment at this stage, in which vocabulary tests and weekly exercises would be adopted in her classroom in order to build up students' capability of learning Chinese language gradually. However, compared with teachers who could adapt well-established test papers, Amber was still not confident to design these tests:

> 所以我想以后慢慢的把这个系统建立起来，而且我自己有时候不知道自己在做，往哪里走。
>
> 在一些 well established subjects 比如 science。每个 unit test 已经决定的。
>
> I want to gradually build this system up, and sometimes I don't know what I am doing and where I am going. In some well-established subjects such as science, each unit test has been decided.

Compared with Diana and Amber, Michelle attempted to integrate summative and formative assessments in improving student learning:

> 我现在越来越觉得，倾向于认为，一个好的 assessment，都是形成性评估。任何好的评估，
>
> 他的目标都是，嗯，都是为了让学生有一个自我评估，同时知道如何进步，就是任何一个好
>
> 的评估， be followed up by feedforward...但是我现在想做的事，比如说，在 end of year
>
> 他是一个 examination,他 assess 的内容就是就是这一整年学的内容。但是，我希望以我的方
>
> 式就是以课后那种有效率的讲评，将把它变成一种新的一种 formative assessment...我会跟
>
> 他们说，你不是被任何一次考试所定义，不管这是多大一次考试....

I believe any good assessments are formative assessment. The goal of any good assessment was, um, give students an opportunity to do self-assessment and know how to progress. That is to say, any good assessment is followed up by feedforward .... What I want to do now, for example, in the end of year, there is an examination, which aims to assess student learning in this whole year. but I hope that I will give effective feedback to students after the examination, so that the examination will be turned into a new type of formative assessment…I will tell them [students] that you are not defined by any test, no matter how important is this exam.

Michelle demonstrated her belief that both formative and summative assessments were opportunities for students to reflect upon learning and make further progress. She reflected that summative assessments were not as effective as formative assessments in promoting student learning, as there was usually no constructive feedback after summative assessment. Based on this understanding, Michelle illustrated that she planned to provide feedback upon each individual student's performance in exams and encouraged them to take opportunities to advance their learning.

Summary

This section has presented the three beginning teachers' conceptions of assessment and their conceptualised practices of formative and summative assessment at Stage 2. Their conceptions regarding the purposes of assessment at this stage revealed that assessment was used to adjust teaching and motivate students to take responsibility for learning. Furthermore, it was found that the beginning teachers conceived that adopting designed assessment activities in the classroom was helpful to use assessment effectively. Nevertheless, the beginning teachers' practices of assessment were hindered by factors such as the teacher's capability of assessment and the assessment context in different secondary schools.

Stage 3: Improvement of Summative Assessment (Term 3 of 2018)

Interviews at Stage 3 took place at the end of the third term of the 2018 academic year. By this time, Diana had left teaching for family reasons and thus her final two observations and interviews were not conducted. Thus, Michelle's and Amber's views of assessment are outlined in the following subsections.

In the interviews, the beginning teachers' assessment experience at this stage highlighted their summative assessment conceptions, given that mid-year exams were happening all around them in the schools at this time.

Purposes of Summative Assessment

By the time the interviews at Stage 3 were carried out, the beginning Chinese-heritage language teachers had implemented some exams in their Chinese classes. Analysis of the interview transcripts showed that they had a focus on two purposes of summative assessment, *demonstrating accountability of teachers* and *improving teaching and learning*. Firstly, Amber and Michelle agreed that teachers illustrated their accountability through students' performances in exams. Amber said that, speaking of assessment, the first idea in her mind was exams, as results of exams went into students' reports. Similarly, Michelle expressed her concerns about taking responsibility for student learning. She asked other colleagues about how much she needed to be responsible for students' achievement, and their responses highlighted the discrepancy between the official response of "trying [her] best in teaching" and the school expectation of "driving excellence." Michelle indicated that although she was teaching Year 9 students and thus felt less stressed than teachers who taught senior students, she was still under the pressure of keeping students learning the subject of Chinese in the following year; if fewer students chose the subject of Chinese, her job might be at risk.

Secondly, Amber and Michelle shared a similar understanding of using summative assessment to inform teaching and assessment. Amber, for instance, demonstrated her conceptions of using exam results to guide practices of teaching and assessment:

> It [exam results] can inform assessment practices like my unit tests. It can inform my teaching practice for next week. It is like a conclusion, like a stop sign, for just wrap up everything, and give me a pretty good idea of where everyone is, and how well they have learnt a certain unit or certain set of learning sequences.

Similarly, but from a different perspective, Michelle illustrated that summative assessment was used for teachers to gain understandings of students, but it had limited influence on improving student learning:

> 是，那么 我看这个 summative assessment 成绩是为了去调整我的教学，去 meet student needs。但是 summative assessment 更有点像是事已至此。不是说不能改变，但是就说到这一个阶段，他是一个什么样的情况，我能够了解这个学生...
>
> Yes, I use [student] achievement in summative assessment to adjust my teaching and meet student needs. But for summative assessment, it was a bit like '事已至此' [what is done is done]. It does not mean that I can't make changes, but when it comes to this

stage, what the students' performances look like, I can gain understandings of this student…

Michelle's conceptions of exams were consistent with her prior understandings at Stage 2, in which formative assessment was used to refine teaching and improve student learning while the summative assessment was used by other stakeholders, such as administrators and parents, to assess achievement of students and make judgements accordingly. At this stage, Michelle pointed out that when teachers used exam results to feed back into student learning, exams could be used as a formative assessment. However, teachers would not normally take further action to use exam results. Michelle planned to make use of exam results by providing detailed feedback to students so that exams would be used to feed back into student learning.

Improving Summative Assessment

During the interviews at Stage 2, both Amber and Michelle had illustrated their plans to refine their practice of summative assessment. Amber had talked about the plan to build up her system of summative assessment, in which exams, unit tests and weekly vocabulary tests were used for motivating students to achieve learning goals gradually. Michelle had planned to provide detailed feedback to individual students after exams, so that students could improve achievement accordingly. In the interviews at Stage 3, these beginning teachers reflected on the meaningful changes they had made in summative assessment practice. Amber shared that she implemented regular unit tests and vocabulary tests in her Year 10 classes:

> I want to know how well they have learnt something, but also equally important, I want them to feel that they have accomplished something, like "I can really speak for a minute or 2 minutes in Mandarin fluently after learning this unit."

Amber indicated the purpose of adopting unit tests and vocabulary tests was to build up students' confidence in speaking Chinese language and motivate them to continue learning. She further explained the reason that she adopted the approach of summative assessment over formative assessment to build up students' confidence:

> Once you tell them, this could go to the report, they sort treat it seriously, if its only formative assessment, they have got lots of other commitments, they may not be as proud of this assessment as other things.

Amber found that students were more motivated to improve their own learning performance when they needed to show accountability in tests. In other words, she believed that employing the approach of summative assessment was more effective in motivating student learning. Amber indicated that she had not enriched her system with exams yet, because she

conceptualised unit tests as important as exams, as every unit test was designed to check students' knowledge; the more that students had learnt in Chinese classes, the more knowledge unit tests would cover. Therefore, once students achieved in unit tests, they were likely to achieve in exams.

Furthermore, Amber still struggled to complete her teaching plans as scheduled at this stage:

> In fact, this year I feel that there is still a lot of content that I have to teach… In fact, we … always can only plan to tomorrow's things. There is not much room for considering future.

Amber explained that she did not implement weekly vocabulary tests as planned before, as she was under pressure to cover teaching content in time.

In contrast, Michelle attempted to use exam results to provide detailed feedback to individual students at this stage. Michelle found that one-on-one oral feedback after exams was effective in catering to student-learning needs:

> I found that when I talk to students one-on-one, the student would feel that he has received more attention, and he would to listen to it [feedback]. That kind of communication will make him feel that the feedback is personalised, and that the teacher is really caring about me.

After each exam, Michelle used one lesson to discuss individual students' exam performance with them. She referenced answers of each student with her marking criteria, discussing the strengths and weakness of their achievement. Then she provided suggestions for improvement to each student. Michelle's reflection on summative assessment was in line with her conception of assessment at Stage 2, in which she stressed that any good assessment is followed by a step of feedforward so that students know their next step for improvement. Michelle reflected that she preferred to talk to individual students after each exam, so that she could discuss students' feedback about exams. However, her school required teachers to provide written feedback to students and their parents, which took lots of her time to prepare. Luckily, her school later changed the policy and gave teachers more flexibility to use exam results to give feedback about student learning, and thus Michelle initiated the one-on-one oral feedback with students to use exam results effectively.

Summary

This section has presented the findings of Michelle and Ambers' conceptions of summative assessment at Stage 3. Compared with prior beliefs of adopting formative assessment practice

at the preservice stage (see Chapter 4), the two beginning teachers showed a shift of emphasis towards using assessment to demonstrate their accountability as teachers. Moreover, these beginning teachers agreed that summative assessment could be used to improve teaching, while Michelle's reflection highlighted the importance of providing feedback to students after exams. Furthermore, factors of the school assessment context and each teacher's capability of teaching and assessment appeared to be influencing their practices of summative assessment at this stage.

Stage 4: Revisiting the Concept of Assessment (Term 4 of 2018)

When the beginning teachers had completed their teaching tasks for the 2018 academic year and started to plan teaching and assessment for the upcoming academic year, they were invited to final individual interviews, to share their stories of learning about assessment throughout their first two years of teaching in New Zealand secondary schools. By the time the last interviews were undertaken, Michelle had renewed her contract with her school for the next year, while Amber had gained permanent employment with her school. Both the beginning teachers had sent in their applications for a full practising certificate.

In the interviews, the two beginning teachers reflected on their changes in conceptions and practices of assessment across the first 2 years of teaching, sharing their gains and challenges in understanding of assessment. The following subsections describe how their conceptions and practices of assessment had changed, from their perspectives.

Purposes of Assessment: An Effective Tool for Improving Teaching

At this stage, the beginning teachers' conceptions of assessment highlighted that their main purpose of using assessment was to improve teaching. For instance, Michelle reflected that she had shifted her conception of assessment from student focused to teacher focused:

> 我觉得以前我就听到 assessment，我觉得它是一个 learner-focused，所有的都是跟学生本身有关，不管是是追求看到一个结果，还是追求为了提高都是针对这个学生。但是我现在对 assessment 的理解会更多的转移到对于教师的 reflection 身上，就是说 assessment 本是教师用于自我学习，调整教学，包括就是说去评估自己跟学生跟课程设计之间关系的一个手段，它更多的把注意力放到教师身上。

> I used to think assessment was learner-focused, as it was related to students, whether they pursue a result or pursue to improve learning. However, my understanding of assessment now shifts to that assessment is the reflection of teachers. That is to say,

assessment is a way for teachers to do self-reflection and adjust teaching. When teachers reflect practices, they will evaluate their relationship with students and curriculum design, thus assessment is more used on teachers' themselves.

Michelle said that she used assessment evidence in the classroom to reflect on the weakness of her teaching and make adjustments accordingly. As an example, she said that she had planned more listening practices for students in the next year due to the fact that students had unsatisfying performances in the final listening exam this year.

Michelle reflected that her changes in conceptions of assessment related to the shifts of the education context and changes in her identity from student to teacher. Firstly, Michelle reflected that she had been studying within the Chinese-heritage culture for 16 years, which shaped her belief that students needed to be responsible for their own learning and thus strengthened her capability of self-reflection. As she began teaching in the New Zealand context, Michelle shifted her conceptions and practices of assessment from the perspective of student to teacher, focusing on using assessment evidence to refine her own teaching practice. Nevertheless, Michelle said that she was sometimes stressed by attributing all problems to herself because if students did not put effort into their own learning, the effectiveness of her refinement in teaching was limited.

Similarly, Amber shifted her conceptions of assessment from the perspective of students to teachers. From the perspective of students, for example, Amber regarded assessment as a tool to check the homework of students. During her first 2 years of teaching, Amber shifted her perspective from student to teacher, considering assessment as a part of teaching, to be used to improve teaching:

> Then assessment is actually designed in every teaching activity, right? Everything we do have assessment. So, I won't think about the problems of assessments, I will think how to improve my way of teaching instead. Why students could not engage in learning? Why can't they say absorb teaching content that I want them to absorb? Because assessment is just a part of teaching activity …so it is every aspect of teaching.

Amber's understanding of assessment was built on the idea that assessment was a part of teaching, as "it always happened in the process of teaching." Based on this understanding, Amber illustrated that she used assessment activities to collect student data and examine the effectiveness of her teaching.

Formative Assessment

In the final interviews, the two beginning teachers reflected their formative assessment practice and their concerns about using formative assessment evidence effectively. Firstly, Michelle and Amber demonstrated different understandings of the concept of formative assessment in the final interviews. Michelle conceptualised formative assessment as an inclusive concept, including any action for students to practise language features and for teachers to improve teaching. She believed that any activity that happened in her classroom was formative assessment, as the purpose of these activities was to assess students' learning and build up their confidence in learning the Chinese language.

In contrast, Amber changed her conceptions of formative assessment distinctively during the first 2 years of teaching. She reflected that she used to conceive of formative assessment activities as the "side benefits" of activities in the classroom. She thought the purpose of her designed activities was for students to review teaching content and have "fun" in learning. Formative assessment only happened when she, the teacher, deliberately collected assessment data in activities. However, Amber admitted that when classroom activities were carried out, data of students' performances in activities comes to her even when she did not collect assessment evidence deliberately.

During the first 2 years of teaching, Amber enriched her understanding through learning in the professional development workshops. Amber shifted her understanding from viewing formative assessment activities as "dispensable games" to considering it as an "essential teaching element" in practice, although she seemed confused about the process of formative assessment:

> 设计,实施,收到 feedback，然后再用。嗯...我感觉他们都是一体的, 是吗？
>
> Plan, implement, receive feedback and use assessment, em…I think they [the four steps] are integrated.

Amber explained that she recognised the four steps of planned formative assessment as an integration, in which each step was used to justify the effectiveness of planned formative assessment:

> 但是如果你设计好了，实施起来就很容易，如果你设计的有问题的话，他们就会坐在那里，
>
> 不知道自己应该干嘛...所以设计跟实施问题是一体的。对，至于 feedback,其实是也是设计的

一部分…比如说老师有精力和空间给学生 feedback，也是因为在设计阶段做的很好，所以老师不用一下来解决这个问题，一下解决那个问题，

But if you planned it well, it will be easy to implement [planned formative assessment]. If you have a problem with the design, they [students] will sit there without knowing what they should do... so the design and implementation issues are integrated. Yes, as for feedback, it is actually part of the design... For example, the teacher has the energy and space to give feedback to the students, that is because she is doing well in the design stage.

Amber thought that the effectiveness of planned formative assessments lay in teachers' preparation at the planning stage. In other words, teachers could predict the implementation of formative assessment activities and students' feedback at the planning stage.

Furthermore, at this stage, Michelle illustrated her struggles to plan formative assessment that was both engaging and effective：

他（学生）选不选课很大程度上就来自于他觉得这个课好不好玩。所以对我们来说，一方面认识到重复其实是语言当中最关键的一个东西，另一方面又要把语言作为一个有趣的选修课要留住学生，所以就不停的往其中加一些这样子(有趣)的活动！

He/she [student] chooses a course or not largely depends on if he/she thinks this class is fun. So, for us, on the one hand, recognising that repetition [practice of language features] is actually the most critical thing in learning languages; on the other hand, we have to shape language learning as an interesting elective course to retain students, so we keep adding some [interesting] activities in the classroom.

Michelle explained that interesting formative assessment activities may be useful in motivating students to continue choosing the subject of Chinese language. Nonetheless, employing various techniques in formative assessment activities may distract students from gaining understandings of their own learning. Michelle said that she discussed her confusion with other language teachers in her school; they all agreed that practising language features repeatedly was an inevitable step in learning languages, students would have fewer opportunities to practise language features when they were attracted by interesting techniques that teachers adopted in formative assessment activities. However, Michelle indicated her inclination to make lessons more interesting rather than providing more practice of Chinese language features due to the pressure of retaining the subject of Chinese. In other words, she

was motivated to engage students, rather than increase the challenge, because if students did not choose Chinese in the next year, her teaching job might be at risk.

In addition, Michelle actively planned various formative assessment activities because she wanted to break the bias from people that teachers from China or Asia teach with mechanical methods. In other words, she was striving to demonstrate that she had a full repertoire of teaching approaches, not just a drill and skill method of teaching.

As regards Amber, she struggled to use formative assessment evidence effectively. Firstly, Amber said that she had a load of student data from their homework and worksheets, but she did not have time to look through their answers and make use of the data:

> 老师太忙，要学生自己对答案，data 就没有达到，到达老师这里就已经消失了。

> When I am busy, I ask students to check answers by themselves, and thus [assessment] data disappeared before it came to me.

Secondly, Amber was struggling to spend more time on refining her practice of formative assessment as she was overwhelmed by deciding "everything":

> 现在的问题是大家什么都有，我们有 word list,甚至还有 language features,里面有 grammar，我们也有自己选择 textbook 的权利，然后我们也有 textbook。但是问题在于，对于新手，那怕是现在的我都没办法 make use of what I have,因为 what I have is too much...

> The problem now is that everyone has everything, we have word list, and even language features, grammar points, we also have the right to choose textbooks, and then we also have textbooks. But the problem is, for the novice teachers, even now I can't make use of what I have, because what I have is too much.

At this stage, Amber was still not confident in making use of assessment data to inform teaching practice. In the final interview, Amber showed her preference for working within the Chinese education system, in which textbooks and exam papers were prepared for teachers. In her opinion, this made teaching manageable, unlike her experience in her first years of teaching Chinese in New Zealand secondary schools.

In fact, as Amber had completed her schooling in both the Chinese and New Zealand contexts, she kept comparing the education systems in her quest to teach and assess effectively. Amber illustrated her preference for working within the Chinese education system, as the effectiveness of teaching was the core of the teaching value:

Yes. 是这样的， 我觉得中国的那些 assessment, 其实我之所以当老师，是因为我觉得 Chinese education 实在太有意思了，我没有办法，把他 get off my head. 对，我其实一直在思考 Chinese way of doing things.我说的 Chinese way includes assessment, 我最喜欢的一点就是他非常的 effective，因为老师们都想 get things done. We want student to learn these things within this period.

Yes, I have been thinking about the Chinese way of doing things. The Chinese way includes assessment. My favourite point is that it [teaching value] is very effective, because teachers want to get things done, we want students to learn these things within this period.

Amber pointed out that, guided by the Chinese-heritage value of pursuing effectiveness of teaching, the system of teaching and assessment in the Chinese context prompts teachers to help students achieve in learning and exams within a certain timeframe. The teaching and research groups in the Chinese context continuously develop and refine the curriculum and exam papers in order to support teachers to a large extent, so that they can focus on refining classroom practice.

However, when Amber began to teach in the New Zealand context, she felt less confident in teaching practice. She explained that the subject of Chinese language was still under construction in her school and in the education system of New Zealand. Amber compared ESOL, that she was also teaching, with the subject of Chinese language, pointing out that she felt less confident in teaching Chinese language, as the subject of Chinese language did not have a solid curriculum in her school:

ESOL 是有至上而下的系统的，第一个就是大纲，他会告诉你七年级必须学完这几个 language features，八年级需要做什么，他把每一年都规划很清楚，就是一个 overall， 大家可以 rely on 的 structure， 这是第一个。在中文教学中，我们现在就有一个 vocabulary list， 里面有 language features. 但是我们需要自己把他 work it out...

ESOL has a top-down system, it has the outline of the course. It shows you things like "you must learn these language features in Year 7," and "what you need to do in Year 8." The outline plans are very clear for every year... Everyone can rely on this outline. This is the first one. In Chinese teaching, we now have a vocabulary list with language features. But we need to work it out.

Amber said that she had to work with other two teachers of Chinese language, deciding the

textbooks and developing the curriculum of the subject of Chinese language. It was overwhelming for a beginning teacher like her. Therefore, Amber wished that there could be teaching and research groups in the New Zealand context, like the ones in China, developing a reliable system for teachers to learn and reference:

> If we have a number of Chinese teachers for teaching many years, they could come up, they could design a really workable system, they can present it to us, we can choose or not, that will be ideal.

In fact, Diana, at her first two interviews in Phase 2, also compared the two education systems in the China and New Zealand contexts (see in this chapter, Stage 1, Summative Assessment Practice), indicating that developing a curriculum and assessment system in a specific classroom context was challenging for Chinese-heritage teachers, given that the teaching and research groups in China provided detailed guidance for teachers and held PD workshops frequently.

Furthermore, Amber reflected that she wished the NCEA system could have a unified textbook for the subject of Chinese language:

> But this subject [Chinese language] is optional. I will definitely design games in order to motivate students to learn Chinese language. So, once I have the textbooks, I will have the idea that what the teaching content is that I have to complete in the worst scenario.

She explained that she compared different textbooks that other schools employed and still struggled to choose the "right" textbooks for students. She wished that there was an official textbook designed by the experts, based on the NCEA standards (as her school uses the NCEA standards), so that she could have a clearer idea of teaching content and put more time into designing formative assessment activities.

Summative Assessment

In the last interviews, Michelle and Amber shared their challenges in refining summative assessment practice. At Stage 2, Amber planned to build a system of summative assessment, in which exams, unit tests and weekly vocabulary tests were considered as checkpoints of student learning. The system aimed to build up students' confidence in learning gradually (Interview 3). At the last stage, Amber indicated that the assessment system for Year 10 students had worked well, as she found the improvement in students' achievement was outstanding. As for the system of summative assessment for senior students, Amber said that

it was challenging for her to build a system to improve senior students' achievement in the NCEA exams within the limited time.

Michelle summarised her changes in practice of assessment during the first 2 years of teaching. In the first year of teaching, Michelle designed tests based on the theories that she had learnt from the ITE programme：

> 我刚出来的时候其实会很怕，就是不都不会，然后所以会抓着理论，就是我从理论出发一定是对的。

> At the beginning of teaching, I was scared as I did not know everything. Thus, I did everything based on theories, as theories never went wrong.

After 2 years of teaching practice, Michelle illustrated that she would design tests based on the specific class context and students rather than learnt theories:

> 现在我先去感受学校的教学环境，以及比如说学校的 summative assessment 的要求，包括那些固定下来的 standards 的，就我们需要测试的 standards,需要 report 的那些东西，然后再根据我课长跟学生的一对一的接触感受他们需要,去调整我的教学。

> Now I prefer to learn more about the school context and the requirements of the school summative assessment first, including those established standards, the standards we need to test and things that we need to be report, then adjust teaching practice based on my one-on-one interaction with individual students.

Michelle pointed out that her teacher education experiences had little influence on her current assessment practice. As she had gained assessment experience in the school context, she had begun to put more thought into designing teaching and assessment under the guidance of the school exam standards, and then adjust classroom practice based on specific class context and students. In this way, she explained that the school's examination standards were a very strong influence on her summative assessment practices.

However, Michelle shared the challenge of interpreting the abstract criteria of school exams in the final interview. She said that it was challenging for her to mark students' performances against the abstract exam-marking criteria for the cultural knowledge standard in her first year of teaching. Unlike the four language skills (listening, speaking, reading and writing) that were assessed in exams in most schools, students' cultural knowledge competence was assessed separately in her school. Michelle had no prior experience in marking exams on

cultural knowledge, and thus struggled to use the four abstract standards of Excellence, Merit, Achieved and Not Achieved to interpret students' performances in exams:

> 他 criteria 也写得非常的模糊，就大概画了一句话，excellent 的就是什 comprehensively，然后 comprehensive 到底是指什么？

> The criteria were also very vague, and there was only a sentence (to explain the criteria). Excellence is students could answer questions comprehensively, then what does "comprehensive" mean?

The abstract nature of the criteria for the cultural knowledge standard prompted Michelle to enrich the marking criteria about cultural knowledge with detailed explanations during her 2 years of practice, although she wished the school could provide detailed criteria for exams on cultural knowledge so that students could know how to better prepare themselves for exams.

Summary

In this section, the two beginning teachers' conceptions about assessment and their experience of formative and summative assessment at Stage 4 has been examined. In general, both beginning teachers demonstrated their emphasis on using assessment to improve teaching, showing that their conceptions of assessment had changed from being more about student-focused accountability, when they began teaching, to becoming more teaching-focused by the end of their 2 provisionally registered years.

Moreover, at the end of the 2 years, these beginning teachers demonstrated different understandings of formative assessment. Michelle considered formative assessment was the aim of classroom activities, while Amber thought formative assessment is an essential part of teaching. Also, while there was insufficient evidence to show that Michelle and Amber had awareness of unplanned formative assessment, Amber's understanding of planned formative assessment showed that she conceived steps of planned formative assessment centred around the planning stage.

Finally, the two beginning Chinese-heritage language teachers mentioned factors that hindered their implementation of assessment in the classroom. Factors of lacking resources and well-established system for teaching Chinese language prevented them from using assessment effectively.

Chapter Summary

This chapter has presented the three beginning Chinese language teachers' conceptions regarding assessment at four stages as they moved toward gaining a full practising certificate

during their first 2 years of teaching. The findings revealed that these beginning teachers' conceptions of assessment changed as they accumulated experience of assessment. Their conceptions in Phase 2 highlighted the purpose of using assessment to improve teaching quality. Furthermore, the beginning teachers demonstrated their conceptions of shaping effective practices of assessment at the different stages of the Phase 2 study, but factors such as school assessment context, their own assessment capability and the school and national systems of assessment shaped the beginning teachers' assessment practices. In addition, it was found that the three beginning teachers shaped their practice of formative assessment as designed classroom activities; how they used formative assessment in the pedagogical process was mentioned less. Therefore, their classroom practice was presented in depth in the next chapter.

Chapter 6 Phase 2: Beginning Chinese-Heritage Language Teachers' Practice of Assessment

Chapter Overview

The previous chapter presented the analysis of the beginning Chinese-heritage language teachers' conceptions of assessment by themes. As noted in Chapter 5, following the Phase 1 focus group interview in 2016 (see details in Chapter 4), Diana, Amber and Michelle agreed to participate in Phase 2 as they began teaching in New Zealand secondary schools. Classroom observations, field notes and reflective conversations were employed to collect data of participants' classroom practice once at each stage throughout the Phase 2 study investigation (see Figure 3.2). After each classroom observation, reflective conversations were conducted at times when the beginning teachers were available. However, as explained in the previous chapter, Diana left teaching for family reasons toward the end of the study in August 2018, and thus the final two observations and reflective conversations with her were not conducted.

In this chapter, the beginning Chinese-heritage language teachers' assessment practices and related conceptions are examined and presented by individual cases. Collected data from interviews and classroom observations was used to understand the assessment practice of these three beginning teachers in different school and classroom contexts. In this chapter, the conceptual framework (see Figure 2.12) was employed as an observation tool to identify and understand the beginning teachers' classroom practice during the observations and analysis. Compared with the interview data in the previous chapter, in which the beginning teachers' conceptualised practice of assessment was presented through the concepts of formative and summative assessment, the observation data from the beginning teachers' classroom practice of assessment was more complex to understand. Thus, the conceptual framework was employed to unpack the beginning teachers' classroom practice of assessment. Also, data from interviews and reflective conversations were adopted as a complement to understand the beginning Chinese-heritage language teachers' assessment practices.

The conceptual framework that guided this study was constructed from the literature as described in Chapter 2. The conceptual framework (see Figure 2.12) included three aspects of teachers' assessment practice to consider: assessment forms (planned and unplanned assessment activities), assessment processes (planning, eliciting, interpreting and using assessment data) and assessment purposes (formative and summative assessment). The

evidence in Chapters 4 and 5 suggests that the beginning Chinese-heritage language teachers mostly did not recognise assessment as a process, and had moved away from it towards more summative, standards-based activity practices over their first 2 years of teaching.

In this chapter, each beginning Chinese-heritage language teacher's assessment practices are organised and presented as two parts: planned and unplanned assessments. This is because planned and unplanned forms of assessment activities were observable and could be captured during my observations of participants' classroom activities. Next, steps of planned assessment (planning, eliciting, interpreting and using assessment) and unplanned assessment (realising, recognising and responding) are presented respectively. Third, data regarding the beginning Chinese-heritage language teachers' formative and summative assessments is presented within each step, as the distinctions between these two purposes became obvious at the stage of using assessment. The chapter concludes with a cross-case comparison of the beginning teachers' assessment practices.

Amber's Practice of Assessment

Amber had her childhood in China. She completed her 9 years of compulsory education in a middle school in China and then moved to New Zealand to continue her study in a New Zealand high school. As a student, Amber experienced assessment innovation within the New Zealand education system. She participated in the NCEA qualification system, a standards-based system for assessment in secondary schools. Thus, Amber had an understanding of the NCEA assessment system used in New Zealand secondary schools as a student. Amber enrolled in the ITE programme in 2016 after she completed her bachelor's degree from a New Zealand university. Before participating in the initial teacher education programme, Amber had no experience in teaching languages. Through learning from the coursework and practica in the teacher education programme, Amber came to embrace different types of assessment and showed interest in using assessment as she began teaching. After graduating from the ITE programme, Amber found a full-time position teaching ESOL and the Chinese language in a Christian school in New Zealand.

The school where Amber was teaching was a state-integrated school[4] that catered for students from Years 1 to 15. The NCEA standards were used to assess the achievement of students. According to the report of the Education Review Office in 2016, the school broadened the

---

[4] State integrated schools are schools with a "special character". This means they may be run by a particular religious faith. Education in state integrated schools is also funded by the government, but they usually charge compulsory fees to help maintain their facilities.

curriculum to meet the various learning needs of students. Chinese language study had been introduced to students from Year 9 to Year 13 to meet learning needs. Amber was teaching both Chinese and ESOL to both junior students (Year 7) and senior students (Years 10 to 13). Her students included both native and non-native Chinese language speakers.

The classroom observations in the Phase 2 study were conducted in Amber's Year 10 classes of Chinese language, while data from the interviews outlined Amber's overall experience of assessment in both ESOL and Chinese language classes.

By the time I first visited her classroom at the beginning of the 2018 academic year, Amber had been teaching for 1 year (2017). As noted in Chapter 5, she had experienced frustration in using assessment effectively and planned to refine her assessment practice in her second year of teaching.

The following subsections summarise Ambers' conceptions of assessment from the previous chapter, followed by details from the four observations of her classroom practice of assessment.

Summary of Amber's Conceptions of Assessment in the Phase 2 Study

The previous chapter examined Amber's conceptions of assessment at four stages as she moved toward her full practising certificate. In the first stage, Amber experienced frustration with assessment and shifted her understanding from viewing assessment as a tool of classroom management towards using assessment to promote teaching and learning. With accumulated experience of assessment, Amber reported that she had become more confident in using formative assessment to adjust her teaching practice in the second stage. Nevertheless, she explained that she was still not confident planning formative assessment activities. The interview at Stage 3 explored Amber's experience of summative assessment and found that she adopted summative assessment to motivate students to take responsibility for learning. At the last stage, Amber espoused the belief that both formative and summative assessment was useful to improve teaching practice. However, restrained by factors such as a lack of resources, Amber still did not feel confident in designing and implementing effective assessment. This analysis suggests that Amber thought of assessment as the first two aspects in Figure 2.11 and did not recognise unplanned assessment opportunities that cropped up within her teaching. Therefore, it seemed important to observe Amber's classroom practice in terms of the conceptual framework (see Figure 2.12), to understand how, if at all, each aspect within the framework was addressed.

Amber's Planned Assessment

Amber's planned assessment activities are considered in this section, including the steps of planning, eliciting, interpreting and using assessment. The following subsections provide details of how these practices were enacted, with examples.

Planning Assessment. As a beginning teacher, Amber was asked to attend a series of workshops as part of meeting the requirements for the full practising certificate. What she found most useful from these workshops was learning that it was important to set up learning objectives for students and show them what success looked like in each assessment event (Interview 2). Communicating learning objectives and success criteria means that allowing students to understand what it is they are expected to learn and what success looks like. Amber demonstrated that she wrote learning objectives and success criteria at the planning stage of planned assessments so that students could follow the right directions and succeed in learning in each lesson (Interview 2).

Amber was observed communicating the learning objectives and the success criteria with her students (Observation 2). However, sharing learning objectives and success criteria mostly happened as she prepared her students for formal examinations rather than in planned classroom assessment activities. For example, Amber showed me a skeleton of her lesson plan before the second classroom observation (see Figure 6.1). She told me that she did not always have time to write down the learning objectives for each lesson, and the learning objectives that she wrote down were only for her use, not for showing in class.

Figure 6.1

Amber's Draft of One Lesson Plan

1. review preposition of places as a class

2. review structures related to preposition of places

3. independent revision (5-10 minutes, ss can choose between quizlet or handwriting)

4. Silent Snap

5. Kahoot: https://play.kahoot.it/#/k/6b4d037a-8bbd-4977-bb55-a20620567c22

6. listening practice

7. Unit 2 writing and speaking assessment explained

8. talk about the marking schedule

9. students work on the speech – drafting

10. homework

As this plan demonstrates, Amber is mostly planning a sequence of activities, rather than clarifying, even for herself, what she intends her students to learn. In fact, only one item in her plan—preposition of places—is a learning objective. The rest of the list denotes *how* she will teach and reinforce these during her class. Thus, the students may not even be aware of what the learning intentions for this lesson were.

Eliciting Assessment Information in Planned Assessment Activities. Amber usually started her lessons by standing at the classroom door and greeting the students in Chinese. From students' brief responses, Amber elicited information regarding students' proficiency in using greeting phases in the Chinese language and then provided corrective feedback to individuals day by day. Amber's collection of student data happened throughout each lesson. At the beginning of lessons, for instance, Amber arranged students to do a vocabulary quiz. Through students' responses in the quiz, Amber elicited data regarding the extent to which students had learnt target vocabulary (Interview 1). Furthermore, Amber often used the strategy of asking questions to confirm if students had understood her instructions. For example, in the first observation, Amber arranged students to repeat zodiac names in Chinese:

Amber leads Ss to read zodiac

A: now it is time for you to read these (four zodiac at a time)

Ss read four at a time with hesitant voices.

Amber repeats then Ss read again. (Observation 1)

Through students' hesitations in reading zodiac names in Chinese, Amber gained the information that students had not mastered the new words. Therefore, she asked students to read the zodiac names again.

Amber employed many methods to collect student information in planned assessment activities. She often collected evidence of students' proficiency at particular language features through peer assessment. For instance, Amber conducted a planned assessment activity to practise dates in Chinese. She selected one student as the questioner, and the questioner needed to ask individual students about their birthday in Chinese. While one student answered the questioner, the others needed to circle down the date in their calendars. After Amber explained the game rules, students started to work on their calendars, and Amber walked around, observing students' work. Amber explained the purpose of walking around later in the second interview:

> There are two choices for me to arrange tasks, I could sit down or circulate, I want to circulate, I want to know what are they doing, I want to know whether someone finds certain things are difficult, if someone made progress, I want to know it. But I will keep walking around them，see if they are doing things they need to do，otherwise，if everyone is behind the device, I am not confident about what they are supposed to be doing, to be honest. (Interview 2)

Amber pointed out her role during the peer assessment activity was to supervise and support students in the planned assessment activities. Through observations of students' performances, Amber gained information on the extent to which students had learnt the target language features.

At the end of the learning period, Amber assessed students' achievement by employing various methods such as role-play, interviews and speeches. Amber indicated that methods of elicitation also influenced students' performances in tests (Interview 3). She preferred to use the method of interviews to assess students' speaking proficiency rather than presentations, as she found that students were easily nervous when standing in front of the class. In line with her espoused beliefs, Amber was observed to provide more flexibility for students to complete the writing test in the second observation. Unlike "normal" tests, which were undertaken in a compact timeframe, this writing test was more like homework; the writing test took place across 2 weeks, as students only had Chinese lessons once a week. Depending on the test topic, students were suggested to draft and then memorise the draft at home. When the writing test

was undertaking in the classroom, students were not allowed to bring their completed draft to the classroom; they needed to write down the memorised draft during class time. Amber explained that the loose-structured writing test reduced the high-stakes nature of assessment for students, and students also had opportunities to practise Chinese language by drafting and memorising the writing test.

The loose-structured tests that Amber employed to collect data on students' performances were related to her prior learning experience in both China and New Zealand contexts. During the interview with the cohort of preservice Chinese language teachers in Phase 1 (2016), Amber mentioned how her 3 years in a New Zealand secondary school had shifted her understanding of assessment. Compared with her experience of taking exams in the Chinese context, Amber felt less stress taking exams in a New Zealand standards-referenced assessment environment, in which "pass or fail" consequences in tests were not as high stakes as in the Chinese context. There, the high-stakes mid-term and end-of-year exams had been critical to her progress through the education system.

Interpreting Assessment Information. It was difficult to observe Amber's interpretation of performances in assessment activities during the four classroom observations. However, before the last observation started, Amber showed how she interpreted students' scores in tests and generated their final reports before the class arrived for the lesson:

> Xx, he did not prepare well, weirdly he had clear logic, but he did not understand a lot during speaking [test], almost Not Achieved, I adjusted his grades to Achieved based on his results of last test. Xx, this is not working, this is Not Achieved. But his oral Chinese was not bad. I will adjust it. Some adjustments I have written on it, for example, this one, he got A and B last time, and Not Achieved this time. I can do some adjustment. If he did not achieve last time, there is nothing I can do. (Reflective conversation 4; translated from Mandarin)

Amber indicated that she did not only use the simplified NCEA standards (see the previous subsection) to evaluate students' performances in the exam, but also referenced their previous performances in the class, even their language-learning environment at home. She indicated that students whose parents intended to have them master Chinese language and created the opportunities for them to speak Chinese in addition to their school experience, had better scores than others in exams.

Thus, with students' two-unit test results and the final exam results in hand, Amber moderated their test results and their learning performance in class, and then produced their end-of-year reports.

Using Assessment Data. Both formative and summative uses of assessment data were found in Amber's observed planned assessment practice: Formative uses included using assessment for student learning and informing teaching practice, while summative uses of assessment data indicated that Amber assessed student performance in exams against her simplified standards. These findings are presented in the following subsections.

Formative Purpose: Assessment for Learning. During the four observations in Amber's classroom, Amber was observed to use assessment data to build up student confidence, provide corrections and advance their learning. First, Amber usually provided confirmation of students' performance, such as "good on you," "well done," or "very good." She also corrected students' learning performance when students made mistakes. In some cases, more practice came after Amber's corrections. For example, Amber provided more practice of reading new words when students' performances were not satisfying:

> A: Nice work, okay let's go over them [words], you have to tell me that you have say it out, on top of?
>
> A: S2, below.
>
> S2: Um… [incorrect pronunciation]
>
> A: 下面，okay?
>
> S2: 下面.
>
> A: Okay, let's not shuffle, 上面… (Observation 2)

Amber prioritised motivating student learning by giving them time and encouragement to refine performances in planned assessments (Interview 2).

Furthermore, Amber provided learning strategies to boost students' improvement in future. After the final examination, for instance, Amber was observed to spend half of her lesson translating test papers line by line (Observation 4). She firstly summarised and explained the most common mistakes during the speaking and listening exam, and then translated the reading material into English. The feedback for improvement emerged from her responses during the planned assessment activities:

A: 我的外公外婆是中国人。S2?

S2: My Grandparents are Chinese.

A: Very good, grandmother and grandfather are different in Chinese actually; we need to learn translate things word to word for NCEA Level 1. (Observation 4)

In the extract above, Amber not only provided the correct answers by translating sentences word by word, but also suggested a learning strategy of direct translation for the NCEA exams in the future.

Formative Purpose: Assessment for Informing Teaching. Amber also actively adjusted her teaching practice based on the assessment data she collected. First, Amber recorded students' performances in classroom assessment activities in order to adjust her teaching plans for the future. For instance, in the last observation, Amber used assessment results in tests to reflect upon the effectiveness of her previous teaching practice. She used half of the lesson to provide feedback on students' performance in exams. She encouraged students to continue improving their language skills, telling them that their performances in the reading and writing exams were less satisfying, maybe because the exam paper was harder than their language levels.

However, Amber expressed her disappointment about some students' results after the class in the reflective conversation. She thought that she had provided enough exercises (such as matching games, board games, etc.) for students to enhance their writing and reading skills, but only three students achieved Excellence. Amber reflected on her own teaching practice and realised that she had not provided enough opportunities for students to practise writing skills, as 80% of class time was used for recognising rather than practising writing Chinese characters:

Because I heard a lot of rumours that if you let the children write, they would be demotivated [in learning Chinese], now they ask for it…but this is a conscious decision, I put priority on interest and communication, did not put in writing. (Reflective Conversation 4)

Amber explained the reason that she reduced the teaching content of Chinese character writing was to motivate students to continue to choose the subject of Chinese language in the next year:

Because they are the only Chinese students for the next year, I need to let them stay interested in this subject, writing practices can wait until next year, as long as they have interests. You can see that they still have needs to practise writing. I am not

155

planning to sacrifice this; I just did not plan to bring it in in this year. In addition, we
do have character writing in each week. (Reflective Conversation 4)

Amber did not prioritise writing skills in her teaching plan this year, as she was under pressure to keep this cohort of Year 10 students choosing Chinese as a subject when they entered Year 11. She had heard from other Chinese language teachers that practising Chinese character writing might demotivate students in learning Chinese language and thus she made her decision to reduce the teaching content of Chinese characters for Year 10 students. Thus, Amber demonstrated how wider pressures also impacted her pedagogy and thus, her opportunities to notice, recognise and respond to students' learning in the classroom.

Summative Purpose: Reporting Students' Performance Against Standards. During the observations, Amber shared her simplified marking schedule (see Figure 6.2) which was used as a standard for generating students' reports (Interview 2).

In this marking schedule, student performance in writing and speaking tests was categorised into four types: Not Achieved, Achieved, Merit and Excellence. Amber explained that this marking schedule was a simplified version of the NCEA language standards. She used to adopt NCEA standards but found that students had difficulties understanding such standards:

Because normally they don't know what they need to do…I would not explain
marking schedule to them before, I thought it was a waste of time, they can look
through it by themselves, it turned out it didn't work out. (Interview 3)

Therefore, Amber simplified the NCEA standards and introduced these to students in the classroom at the second stage. From students' responses and their interactions about this marking schedule, Amber believed students had figured out what they needed to do to achieve in the writing and speaking tests. Also, Amber pointed out that adapting NCEA standards was helpful for "building up for the future who would take NCEA" (Interview 3).

Figure 6.2

Marking Schedule for Speaking and Writing Tests

2018 Year 10 Chinese Unit 2 Writing and Speaking Assessment

Student name:

Marking Schedule

Components	NOT ACHIEVED	ACHIEVEMENT	MERIT	EXCELLENCE
Writing	Task not attempted OR Covered one aspect only. Did not always select and use language features that are fit for the purpose and audience. Meaning is often obscure because of inaccuracies.	Covered two or more aspects. Selected and used language features that are generally fit for the purpose and audience. Communication is achieved overall despite inaccuracies such as grammar errors and word choice.	Covered all three aspects. Ideas are developed, writing is generally controlled. Selected and used language features that are mostly fit for the purpose and audience. Inaccuracies do not significantly hinder communication.	Covered all three aspects with considerable details. Ideas are well developed, writing is controlled. The student capably selected and skilfully used language features that are fit for the purpose and audience. Inaccuracies do not hinder communication.
Speaking	Task not attempted OR Speech is less than one minute in length. The student frequently referred to cue cards. Meaning is obscure because of inconsistencies in pronunciation and intonation.	The speech is at least one minute long. The student sometimes referred to cue cards. Demonstrated some degree of confidence and fluency. Communication is achieved overall despite inconsistencies in pronunciation and intonation.	The speech is one to three minute in length. The student occasionally referred to cue cards. Demonstrated increasing confidence and fluency. Communication is not significantly hindered by inconsistencies in pronunciation and intonation.	The speech is one to three minute in length. The student seldom referred to cue cards. Demonstrated fluency and confidence. Communication is not hindered by inconsistencies.

Amber's Unplanned Assessment

As examined in Chapter 2 (see Figure 2.12), in addition to planned assessment activities, unplanned assessment opportunities exist within teaching and learning situations. Thus, it was important to observe, as much as possible, both the planned and unplanned assessment activities, including those with individual students and small groups. In line with this

understanding, unplanned assessment in this study was observed to be often embedded in teaching or even to occur within planned assessment activities.

During the classroom observations, Amber was observed employing formative assessment when students asked questions or when she noticed group discussions surfacing learning issues and challenges. Also, Amber was also observed to ask students questions when they were not participating in assigned tasks. The information that Amber received was used to encourage students in learning, help them construct new knowledge, engage students in learning and inform her teaching practice.

Firstly, Amber was observed encouraging student learning by specifying their attainment when she noticed students were challenged. For instance, in the final observation, when a student told her the way to memorise the words with similar character forms, Amber confirmed his method of memorising Chinese characters by specifying his attainment:

A: Um… I think you need to do a bit more than that.

S: I know how to write 外公外婆, that is two different characters than 妈妈.

A: That is a great way of putting it. So 爷爷奶奶, same character repeated twice, is grandpa and grandmother, 外公外婆 different characters are maternal. (Observation 4)

Secondly, Amber used assessment information to help her students build their knowledge. For example, when Amber interacted with students during the second observation, she received questions from students and responded to students with corrections of different translations of "grandparents" in Chinese language and then further constructed their knowledge regarding the differences between "爷爷奶奶" and "外公外婆." Similarly, Amber got used to constructing students' knowledge after recognising their questions in the classroom:

S: What does second character mean?

A: 在，there is more formal translation for that. Thank you for the question, it means at the position. I decided not to explain too much, just giving you the structure, A 在 B 右边。But that is a great question, thank you for asking. (Observation 2)

In this extract, this unplanned assessment was embedded in a planned assessment activity. Amber asked students to translate Chinese into English for her. After the corrective feedback

of "猫在车子上," a student further asked her to explain what the preposition "在" was. At this moment, Amber noticed that students had not learnt how to use the preposition 在. Then she provided more instructions to all students regarding how to use 在 in sentences.

Another example happened when Amber was preparing her students for the upcoming test during the third lesson observation. While Amber was explaining the marking schedule for the class, one student asked the meaning of "inaccuracy" in the marking schedule. Amber explained the meaning of this word, and then she intentionally responded by using this word in the following instructions to deepen students' understandings of inaccuracy.

Thirdly, Amber was observed noticing student behaviours associated with competence, and responding appropriately to increase the review activity. In this example, during my second lesson observation, when Amber walked around and supervised students to do a vocabulary review activity, she noticed that several students had finished their arranged tasks and started talking. She recognised students with high levels of Chinese language had completed the assigned work, and then she responded by increasing the challenge for this group of students:

> A: Do you have A? Wow…keep going, just keep going, okay? I only want your best time. (Observation 2)

In the extract above, Amber adjusted the assessment activity for high achievers in order to engage them in different levels of the activity.

Finally, Amber was observed adjusting her teaching practice in response to noticing something. For example, in the second observation, Amber "paused" a little after a review activity of vocabulary; she asked them if they still needed time to review the vocabulary. With no surprise, students called out that they did not need more time to do the review. Thus, they moved to the next activity. After the class, she explained the moment of pausing:

> [pointing to teaching plan] There is no match game, I didn't know that they have learnt this part and performed well, I thought they may still struggle to speak out or make mistakes. So, I was planning to give them 5 to 10 minutes to do independent revision and then do the salient snap. It turned out they were great on this part, so we spent 5 minutes to play a matching game. (Reflective Conversation 2; translated from Mandarin)

In the extract above, Amber replaced the planned independent revision with a "matching game," as she noticed students had mastered the work that she assigned. Thus, she changed the activity based on students' learning outcomes.

Summary of Amber's Practice of Assessment

This section has provided descriptions of Amber's practice of assessment during her 2 years of provisional registration as a teacher. Amber changed her conceptions and practices of assessment during the Phase 2 study. In the interview during Stage 1, Amber claimed that she did not value students' feedback in class, as "they could not provide correct answers." However, the analysis of Amber's classroom practice of assessment showed that Amber actively used both formative (planned and unplanned) and summative assessment to build student achievement, engage students in classroom learning and adjust her teaching practice.

Moreover, in contrast with her statement in the final interview, in which formative assessment activities happened only when she planned deliberately, Amber was observed frequently using both planned and unplanned formative assessment activities to assess student-learning performances. The inconsistency between her espoused beliefs and actual practices showed that Amber did not recognise her use of formative assessment as part of her teaching. She appeared to understand formative assessment as planned events without recognising that her ability to notice, recognise and respond to students in teaching processes was a practice of formative assessment.

Michelle's Practice of Assessment

Michelle completed her schooling in China and had gained her master's degree in Chinese literature at one of the top universities in China. During her study at the university, she received an opportunity to go to New Zealand and support local primary and secondary schools as a Mandarin language assistant (MLA) for 1 year. During her 1-year practice, Michelle gained experience and interest in teaching Chinese language in New Zealand secondary schools, and thus she enrolled in the ITE programme and launched her teaching career in a New Zealand private school after graduation.

The school where Michelle worked is an independent school[5], catering for students from Year 1 to Year 13. This school offers both NCEA and IB qualifications for Year 12 and 13 to choose from. Michelle was teaching the subjects of Chinese language and social studies for

---

[5] Independent schools (Private schools) are not directly controlled by the government. They must meet certain standards to be registered as a school but are governed by an independent board. Independent schools receive some funding from the government but are mostly funded by yearly school fees and donations.

students in Year 7, Year 9 and Year 12. Furthermore, Michelle was not only teaching but was also responsible for developing the curriculum for the course on Chinese literature.

In Phase 2, the observations were carried out in Michelle's Year 9 classes of Chinese language in order to record Michelle's practices of assessment, while her self-reported experience of assessment in the interviews covered her practices in different subjects across her Year 7, 9 and 12 classes.

During the four classroom observations in Phase 2, as with Amber, both planned and unplanned assessment activities were noticed in Michelle' classroom practice. The following subsections firstly summarise Michelle's assessment conceptions of assessment in Phase 2, then present her classroom practice of assessment with examples.

Summary of Michelle's Conceptions of Assessment

During the focus group interview in Phase 1, Michelle shared her experience about learning assessment in different contexts and her passion to integrate both formative and summative assessment into the classroom when she started working in New Zealand schools.

At the end of 2017, Michelle was invited to the first Phase 2 interview. She shared both her gains in understanding and using assessment in the first year of teaching and the pressure from the school which asked her to apply for a visa for permanent residence to secure her contract with the school for the next year. By the beginning of Stage 2 (February 2018), Michelle reported that she had found an approach to delivering teaching and assessment in the classroom. She had built up a database for planned assessment activities including both easy and difficult activities to assess student learning, although she reported that creating these resources for planned assessment activities was "time-consuming" (Interview 2). At Stage 3, influenced by changes in school policy, Michelle reported that she had improved her feedback practice based on students' test results by showing individual students how to improve.

After 2 years of teaching, Michelle reflected on her teaching experience in the final interview. On the one hand, Michelle felt confident in using assessment, as she had built up a database of formative assessment activities and found a way to use exam results to provide feedback on student learning. On the other hand, Michelle described her experience of both teaching and assessment as "战战兢兢，如履薄冰" (skating on thin ice). She kept feeling the pressure of retaining her job, as she had not gained a permanent residence visa. Moreover, it was still challenging for Michelle to create assessment resources (such as formative assessment

activities and exam papers) when lacking support from the teacher in charge of the Chinese language subject.

In the following sections, Michelle's actual classroom practices are examined under the subheadings of planned and unplanned assessment, in which Michelle's espoused beliefs related to actual practice are presented to understand her classroom practice.
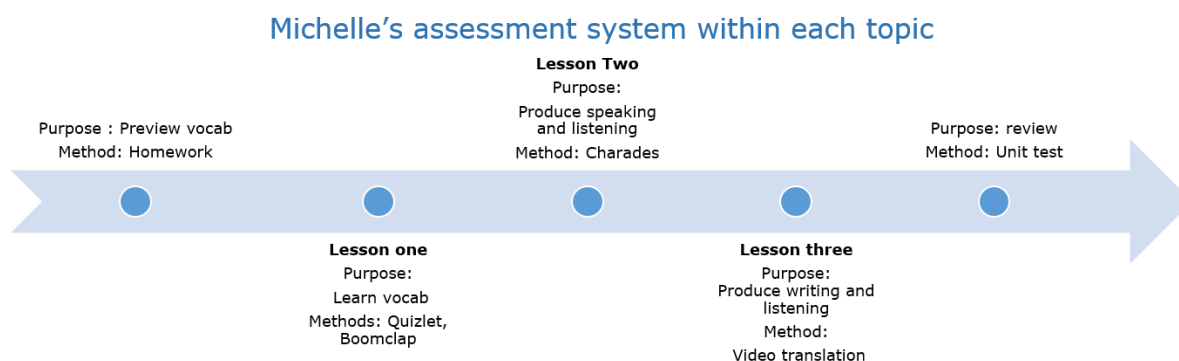
Michelle's Planned Assessment

During observations of Michelle's classes, her practices regarding planned assessment activities are examined through the lens of the conceptual framework (see Figure 2.11).

Planning Assessment. Michelle showed her confidence in planning assessment activities throughout the data-collection procedure. She stated that she used most of her time to design various classroom assessment activities (Interview 2). Before the second classroom observation, Michelle had built up a routine of formative assessment activities for each learning topic (see Figure 6.3). Michelle explained that the purposes of using this routine was to help students and herself remember the objectives and check up with planned assessment activities at different points of each language topic. For example, Quizlet (a web-based vocabulary learning application) was mainly used in the first lessons of each topic to assess students' recognition and pronunciation of new words, and a game of charades was used in the second lessons to assess students' speaking and listening skills. In the last lesson of each topic, Michelle played videos and asked students to write down video scripts, so that she could assess students' mastery of listening and writing skills. Michelle reflected that it was helpful to develop this routine of formative assessment activities. Furthermore, she mentioned that she prepared two sets of formative assessment activities in this routine, with easy and difficult levels, so that she could adjust the degree of difficulty to meet student needs. With the routine of formative assessment activities, Michelle stated she was flexible and confident to implement and refine assessment activities within this routine (Interview 4). Also, both she and her students could easily trace their learning through this routine of planned assessment activities. Michelle could further make adjustments when students' performances were less than satisfying. She said:

> Students will have the idea what game will be played at the beginning of each topic, and in the last lesson of this topic, they know they will watch video and write down video transcript. (Reflective Conversation 2; translated from Mandarin)

162

Figure 6.3

Michelle's Routine of Formative Assessment Activities Within Each Topic

Eliciting Assessment Information. As introduced above, Michelle built up a routine of planned assessment activities for collecting information about students' proficiency in Chinese language at different stages of each topic. At the beginning of lessons, Michelle was observed to employ the approach of both formative assessment (such as Quizlet and brainstorming) and rewarding achievement (such as giving stamps for students who had finished homework). She also collected evidence of student learning from these activities, thus meeting a summative purpose. For instance, Michelle employed a web-based application called Quizlet to assess students' proficiency with new words. In the second observation, when students entered the classroom, they were arranged randomly into groups and answered questions generated by the Quizlet. Michelle shared that she normally did not circulate in the classroom and provide help during this process, but students would easily settle down at the beginning of lessons by being engaged in interesting classroom activities. Also, they had opportunities to assess and review their learning during the activity (Interview 3). After the Quizlet activity, Michelle discussed the content of this assessment with students:

> M: Very good, find your seat and take out your notebook, I have a question, why do you think today's Quizlet live is getting a bit trickier?
>
> S: As we have a new topic.
>
> M: We do have a new topic.
>
> S: We have learnt some new words.
>
> M: We have learnt some new words, we were not that familiar with, because that is what we focusing on. But I am aware some of you have started the Language Perfect, which is great, we can have a look. (Observation 2)

In the extract above, Michelle encouraged students to find out the factors which may influence their performances. Students pointed out that the Quizlet activity was harder than before as there were words in Quizlet that they had not learnt yet. Then Michelle praised students who had previewed vocabulary by doing the assigned homework on the Language Perfect (an online learning system). After the lesson, Michelle further explained that she had predicted that students might have trouble in understanding and using new vocabulary. This was because she had checked students' completion of assigned homework for the week and found that only a few students had completed or started doing homework (Reflective Conversation 2).

Furthermore, Michelle also adopted the approach of summative assessment to collect information about student learning. In the second classroom observation, Michelle and her teaching assistant were observed checking students' homework and giving stamps at the beginning of the lesson. Michelle claimed that the purpose of giving stamps was to motivate student learning, as she planned to give prizes to those who had enough stamps at the end of each semester (Interview 2). However, as discussed in the final interview, Michelle found the stamp system did not work well for Year 9 students, as they had enough motivation to learn Chinese language at home; completing homework for them was routine.

During lessons on each topic, Michelle was observed assessing students' proficiency in Chinese language through frequent planned assessment activities (Observations 1 and 2). For example, in the second classroom observation, Michelle used the game charades to assess students' mastery of speaking about places in the Chinese language. One student acted in the front and the others needed to answer Michelle's questions regarding "Where is she going?" in the Chinese language:

> [One student is acting]
>
> M: Where is she going?
>
> Ss: Sports field.
>
> M: Yes, sports field, ok we all identify it is sports field, everyone look at the board.
>
> Michelle showed the correct answer in the video, the voice from video said 运动场 (Chinese for sports field).
>
> M: Repeat after it.
>
> Ss: 运动场 运动场. (Observation 2)

As shown in the extract above, Michelle elicited students' proficiency in speaking place names in Chinese by observing their responses to one student's performance on the stage. Through students' responses in the English language, Michelle realised that students could understand the meaning of the performance, but it was still difficult for them to pronounce answers in the Chinese language. Thus, Michelle added further practice, playing the correct answer in the video and asking them to repeat the answer so that students had more opportunities to practise tones and memorise the meaning of the character.

At the end of each learning unit, Michelle adopted the approach of unit tests to assess students' learning in the topics. However, as Michelle said in the third interview, she was responsible for designing eight school-required progression exams, unit tests, at the end of each learning topic that were generated by the online learning system Language Perfect, which included tests of listening, reading and writing. Students could see their test results directly after the tests. If their scores were under 70 points, the Language Perfect system would assign further practice for students. As Michelle illustrated in the third interview, she paid attention to students who had not achieved 70 points, checking their tests and considering if she needed to provide further instruction for those students.

Using Assessment Data. After collecting and interpreting assessment data, Michelle used assessment information in various ways. In this section, themes of formative use of assessment data on student learning (Observations 1, 2, and 4), and informing teaching practice (Observation 4) and reporting student performances (Reflective Conversation 4) are examined and presented.

Formative Purpose: Assessment for Learning. During the four classroom observations, Michelle used assessment data to provide corrections, build up student confidence in learning, construct new knowledge and advance student future learning within formative assessments. Michelle's classroom practice was in line with her espoused beliefs about assessment, which was any good assessment should be formative assessment (Interview 2).

First, Michelle's corrections focused on students' performances using features of Chinese language. For instance, in the second observation:

M: 假期你去哪儿？ (Where have you been in the holiday?)

S: 我住在奥克兰。 (I live in Auckland)

M: You can say "我在奥克兰"。 (You can say "I stay in Auckland")

S: 我在奥克兰。 (Observation 2)

Michelle provided corrections directly when a student was having difficulty answering her question in Chinese language.

Furthermore, assessment information was used for "care-referenced" purposes. Michelle was observed to encourage students in order to build up students' confidence in learning Chinese language throughout the four classroom observations. For example:

M: Lastly, who can read it? Xx.

S: 他们是英国人吗？ (Are they British?)

M: 很好的 translation, 他们是英国人吗？ (The translation was good, are they British?). (Observation 1).

Michelle provided specific praise for the student's translation skill and repeated her translation. During the classroom observations, Michelle was observed to praise students' performances in the Chinese language by repeating their answers. Furthermore, Michelle encouraged students by inquiring about the students' methods of learning Chinese characters. For instance:

S: Is it like a gym?

M: Very good, how do you recognise it?

S: As it's Pinyin.

M: Very good, it's like PE. (Observation 2)

Michelle confirmed the student's answer and further inquired about the method that he used to recognise the new Chinese character for "gym." After the student explained that he recognised the Chinese characters by Pinyin (the Romanisation of the Chinese characters based on their pronunciation), Michelle praised the student again.

Third, Michelle was observed to use assessment data to construct students' new knowledge. During the classroom observations, after Michelle received assessment information from students in planned formative assessment activities, she usually prompted students to solve problems themselves rather than providing correct answers (Observations 1, 2 and 4). For instance, during the first lesson observation, Michelle played a video clip and asked students to translate the dialogue into Chinese. After students called out their answers, Michelle did not provide the correct answer right away. Instead, she decided to play the video again:

M: Try to be very accurate in terms of "this" or "that," listen again; I want some volunteers to translate. (Observation 1)

166

By reminding students to be accurate in translating "this" and "that" into Chinese before she played the video again, Michelle gave students had another chance to refine their answers. In fact, during the lesson observations, when students sought corrective answers from Michelle, she barely provided correct answers to students directly. Instead, she asked students further questions to help them find the answers themselves (Observations 1 and 2). For instance, in the first observation, Michelle constructed students' success in speaking the Chinese language by breaking up her question and then inspiring students to translate word by word:

M: Did you notice that sometimes we say "我妈妈," sometimes we say "我的妈妈." Does it make any difference?

Ss: Yes.

M: S1, can you explain it?

S1: I don't really know; they are just different.

M: S2

S2: There is a difference between 我的 and 我。

M: Close, close, S2?

S3: You say when you have said the top sentence like that?

M: So here is the thing, the formal way to say it is "我的妈妈" …我 means….

Ss: I.

M: 我的 means…

Ss: My.

M: So when you say my mum, you say "我的妈妈" … [teaching instructions].
(Observation 1)

Michelle did not provide explanation directly. Instead, she led the students to discuss the differences between "我" and "我的" based on her answers.

Finally, Michelle was observed using assessment data to advance student learning in future. In the last lesson observation, Michelle gave feedback to students about their writing exam results. She first gave 10 minutes for students to correct answers in their exam papers. Michelle told students that they could either correct the exam papers by themselves or ask for peer help. When Michelle noticed that students had almost finished their corrections, she

asked individual students to come up to her desk and discuss their performance in the exam. As Michelle reflected after the lesson, it was an opportunity for both students and her to ponder their Chinese achievement in the past academic year. Students could ask questions about the exam during the individual feedback, and she could have an opportunity to summarise their performance this year and encourage them to continue learning Chinese.

Formative Purpose: Informing Teaching. Michelle was observed using data collected from planned assessment activities to adjust her instruction. For instance, during the second observation, Michelle gained assessment information about students who could not pronounce the correct answer to the word "运动场" (sports field) when few students replied to her question. Thus, she tried a new way to read the word in the activity and asked students to repeat after her:

> M: Ok, everyone, I want you to notice your clarity of pronunciations, especially for this one, listen, and repeat two times. (Observation 2)

In fact, as shown above (see Figure 6.3), Michelle employed different planned assessment activities (such as Quizlet and charades) to assess student learning continuously at different stages of the topic. These planned assessment activities were interwoven with the teaching instructions, guiding Michelle to adjust teaching time with flexibility:

> When I was circulating and checking their performances during assessment activities, I can see their participation in the activity. In terms of the exercises of reading and speaking, I got information of student learning from observing their performances in practice. In addition, I can see the effectiveness of these activities from students' performances in the planned assessment activities in the next lesson. My teaching plan for the next lesson will be as usual, but the teaching time will be flexible. (Reflective conversation 2; translated from Mandarin)

Michelle considered planned assessment activities as a way to assess students' learning gradually throughout the unit. Her teaching plans in the next lesson were flexible enough to respond to students' performances demonstrated through the planned assessment activities.

Summative Purpose: Reporting Student Performance in the Written Exam. Michelle reflected on her practice of reporting student performances in the final writing exam during our last reflective conversation. She explained that she adopted the languages standards in their school to design exam papers, as shown in the Table 6.1.

Table 6.1

Marking Standards for Writing Exams: Michelle's Assessment Practice

Achievement	Achievement with Merit	Achievement with Excellence
At least 90 Chinese characters are written	At least 90 Chinese characters are written	At least 90 Chinese characters are written
My script is clear and easy to understand	My script is clear and easy to understand	My script is clear and easy to understand
I write on the topic in Chinese	I write on the topic in Chinese and give some extra details to make my writing more interesting	I write on the topic in Chinese and give a lot of extra details to make my writing more interesting
My words and sentences make sense some of the time. I can communicate some of my ideas	My words and sentences make sense most of the time	My words and sentences make sense all or nearly all of the time
I use basic words and structures including what we have just learnt	I use a variety of words and structures including what we have just learnt	I use a wide variety of words and structures including what we have just learnt

Michelle explained that there were three standards for the writing exams: Achievement, Achievement with Merit and Achievement with Excellence. After she collected students' exam papers, Michelle would ask her teaching assistant to circle students' mistakes regarding grammar and vocabulary, and she would check exam papers for mistakes in sentence structure and logic. Then, Michelle and her assistant would moderate students' scores in the exam. They discussed each student's performance against the marking standards and categorised each student's performances into a suitable standard. Finally, Michelle entered students' performances in the exam into their end-of-year reports, as she commented:

> 其实 report 是给家长看的，但是同时学生他们其实自己也会看。report 他之所以需要写，其实也是为了让学生和家长对学生的学习情况有数。所以report其实是想怎么说，这个考试是为了让学生心里有数儿，其实是与之而来的一个行政文件。

> In fact, reports are for parents. But at the same time, the students will actually read it themselves. The purpose is actually to let students and parents know the student's

learning situation. So, the report wants students to know about their learning situation, and it is also an administrative document. (Interview 3)

Michelle's Unplanned Assessment

During classroom observations, unplanned assessment was observed as interactive formative assessment embedded in Michelle's teaching instructions and planned formative assessment activities. Unplanned assessment was triggered by students' questions or Michelle's observations. After noticing and recognising students' performances, Michelle was observed using this information as assessment data to give feedback or adjust teaching and assessment practice. For instance, during the second visit to Michelle's classroom, students were arranged in pairs to play a "match game," guessing Chinese words on each other's paper. While the matching game was ongoing, Michelle walked around and noticed one student was having trouble finding words on her sheet:

> M: You haven't played this before, right?
>
> S: No.
>
> M: So xx, can you ask S a question in your sheet?
>
> [M continues instruct both of students]. (Observation 2)

Michelle recognised assessment information by confirming whether the student had played this game before. Then she responded with detailed instructions to students.

Apart from responding to students' questions directly, Michelle motivated students to explore answers to their own questions:

> S: How does that show your cultural aspect?
>
> M: What do you think? It has to include everything in Chinese, but how does that demonstrate your cultural understanding? How does that differ from a writing assessment, which is on my school life?
>
> S: As you are creating.
>
> M: This is the criteria that we used in other languages, so it is kind of broad. But how do you demonstrate your cultural understanding through a booklet of your school life. How does that differ from a piece of writing? What do you think? You have time to work on that, and you have an oral presentation of course, it is creative, but in the content, what makes it cultural?

S: …

M: Maybe the cultural community of the school, which nationality are they in the school? And like the people you know and your friends, who they are, and who you engage with throughout the day. Sounds good? Anything else about cultural assessment? (Observation 3)

When the student asked Michelle the question about how to demonstrate their cultural knowledge in their homework, she did not respond to the student directly. Instead, Michelle further asked all the students detailed questions to inspire students to unpack the question. When she noticed that students still had difficulty, she provided her understanding of the questions.

The observation data were in line with Michelle' s conceptions of assessment, in which students played the central role during classroom assessment; she was only active when she needed to explain things to students (Interview 3). When students focused on completing their tasks, Michelle was observed to circulate in the classroom, observing students' performances and providing support to individual students (Observations 1, 2, and 4).

Moreover, Michelle was observed using assessment data to adjust her instructions or assessment practices instantly. During the second classroom observation, Michelle intended to ask students to participate in a game of charades, in which students' listening and speaking skills were assessed. Before the game started, Michelle had led students to practise useful words in the game. When she asked students if they were ready for charades, many students whispered that they were still not ready:

M: Okay, do you think you are able to do the charades, as you will only be reading Chinese characters?

Ss: No….

M: Okay we will quickly flick through it, let's just call out the meaning. (Observation 2)

When Michelle received assessment information that students still needed practice, she adjusted the activity by adding more translation practice before the next planned game.

Summary of Michelle's Practice of Assessment

In this section, Michelle's classroom practice of assessment has been examined. Michelle appeared to have a clear understanding of integrating both formative and summative

assessment into her classroom practice. During the classroom observations, Michelle had realised her goal by building up the routine of teaching and assessment practice (See Figure 6.3), in which both formative and summative assessment approaches were used to motivate students to improve gradually. Michelle's success in building up a database of teaching and assessment within her first 2 years of provisional registration was related to her Chinese cultural background. Michelle explained that the Chinese culture of examinations strengthened her self-reflectiveness from her identity as a student to a teacher, like she explained in the interview at Stage 4:

> 所以其实我从国内到这边的一个 shift，他其实加强了我的 self-reflectiveness，但是是因为我的身份也在跟着 shift，我的环境也在跟着 shift。因为我在国内的时候，我的性格慢慢形成了一个性格，是会把什么都 attribute 到自己身上，就是我如果考得不好，我不会讲是老师是不是这个就没交清楚，我什么都会觉得是我学的还不够，所以我会更加的努力，自己去补全自己知识的漏洞...到了这边来说，因为我变成了老师...在这边学习到的对于 assessment 的认知提高了，就把它变成了跟老师有关的，所以它其实加强了我对自己的反省。

> In fact, I shifted and strengthened my self-reflectiveness when I came to here [New Zealand], because my identity shifted, and my environment also shifted. When I was in China, I gradually shaped my characteristics, which is that I would attribute everything to myself. That is, if I failed the test, I wouldn't think it was the teacher's fault, I would reflect that I haven't learned enough, so I would work harder to fill in the gaps in my knowledge…When I came to here [New Zealand], because I have become a teacher…I have learnt a lot about assessment, I turned to think about assessment from the teacher's perspective. So, it actually strengthened my reflection on myself. (Interview 4)

Diana's Practice of Assessment

In this section, Diana's classroom practice of assessment during most of her 2 years of provisional registration are examined. Before Diana came to New Zealand, she completed all her education in China and was teaching ESOL for 11 years in a Chinese secondary school. The school where she worked was in Beijing, where the competition among students and schools was fierce. In China, Diana was teaching English language to students from Year 11 to Year 13, preparing them for university entrance examinations. Thus, she claimed that she

was under stress in her teaching job on the one hand, while, on the other hand, she honed her skills of teaching by attending many workshops and meetings.

Diana and her family moved to New Zealand, leaving her teaching job in Beijing. She enrolled in the ITE programme in order to gain New Zealand registered-teacher status and continue her teaching career in New Zealand secondary schools. In Phase 1, Diana illustrated her changes on conceptions and practice of assessment from the high-stakes assessment environment in China to an evidence-based assessment context in New Zealand. After 1 year in the ITE programme, Diana successfully graduated from the ITE programme and found a part-time teaching position in a private school in New Zealand. She showed interest in trying both formative and summative assessment to improve learning and teaching as she began teaching.

The private school she worked in catered for students from Year 1 to Year 13. It offered both the CAIE system and NCEA qualifications; students had options to plan their learning and assessment within these different qualification systems. Diana was teaching ESOL and Chinese language for Year 7 and Year 10 students, and she also undertook tutorials for students and teaching tasks for short-term study programmes in her school.

In Phase 2, Diana reflected in the interviews on her experience of assessment when working with both junior and senior students, while the classroom observations were carried out in her Year 10 Chinese language classes. However, Diana left her teaching job for family reasons during late 2018, and thus the last two data-collection sessions were not conducted.

In the following subsections, Diana's conceptions of assessment in the Phase 2 interviews are first summarised. Then her classroom practice of planned and unplanned assessment is presented with details.

Summary of Diana's Conceptions of Assessment

Diana's conceptions of assessment related to her experience of adapting her teaching experience in China into the New Zealand context. At the first stage, Diana reported increased confidence in planning formative assessment activities as she had accumulated classroom practice, although she commented that she lacked resources and had insufficient communication with counterparts and that these issues hindered her practice of assessment. Furthermore, Diana placed school exams as the focus of her practice because she saw meeting the requirements of school assessment as paramount (Interview 1). At the second stage, Diana stated that she had become more experienced in planning and implementing both formative

and summative assessment although she said that her emphasis in practice was still summative assessment. This was because both teachers and students in her school were under much pressure to achieve in exams, and Diana wished to demonstrate her accountability to students, parents and the school (Interview 2).

During the classroom observations, both planned and unplanned assessment was observed, and the evidence is presented in the following sections.

Diana's Planned Assessment

During classroom observations, Diana's classroom assessment practice was recorded and demonstrated different strategies at each stage.

Planning Assessment. During my visits to Diana's classroom, she reflected that she had a tight schedule to complete teaching content (see the course outline in Appendix H). Therefore, she cut down the number of formative assessment activities in favour of covering teaching content in each lesson (Interview 1). Interestingly, While I was observing her class, the headings "Lesson Plan," "Learning Objectives," "Success Criteria," "Key Words," and "Homework" were listed on the whiteboard in her classroom. Diana explained that she was required by her school to introduce these to students before each lesson. She found this strategy was useful as students and she could have a clear picture of what would happen in each lesson (Reflective Conversation 1). However, during the first classroom observation, Diana read through the content on the whiteboard and had little communication with students about the learning objectives and success criteria in this lesson.

Eliciting Assessment Information. Diana was observed to conduct assessment activities to elicit students' proficiency of the Chinese language. The results of these activities were used for both formative and summative purposes. As well as conducting unit tests and exams, Diana designed various assessment activities to engage students in learning. At the beginning of each lesson, Diana used a few minutes to ask students to do speaking exercises. From these exercises, Diana received information regarding the extent to which students had learnt. For instance, at the beginning of the first classroom observation, Diana elicited students' proficiency in vocabulary by asking them to do translation:

D: What is "can/may," in your booklet, vocab?

S1: 可以

D: 可以，right, that is what we will do [writes the character on the board], S2, give.

174

S2: 给

D: 给, very good, S3, make a phone call.

S3: 打电话. (Observation 1)

During lessons, Diana employed many methods, such as questions and answers, practice dialogues in pairs, and competitions in different groups, to elicit students' language performances. For instance, in the second lesson observation, after learning Chinese words for food, Diana designed a mini dialogue regarding ordering food in the cafeteria. She asked students to practise in pairs and then act it out. It was useful to find out the extent to which students had learnt the new words.

At the end of each learning period, Diana conducted unit tests, progress tests and examinations to record students' performances of learning (see Figure 6.4).

Figure 6.4

Diana's School Assessment Structure for Year 10 Students
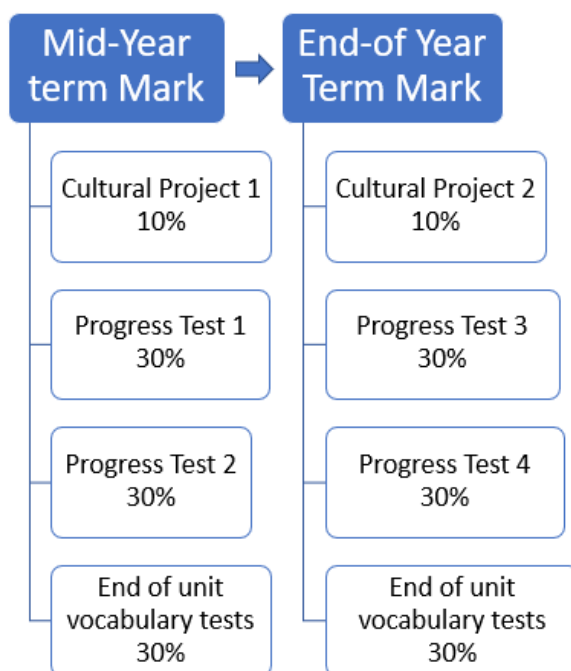
In Diana's school, Year 10 students' performances were assessed and marked at mid-year and end of year. In these, students' scores in cultural projects, progress tests and unit tests were recorded for the end-of-year academic excellence awards in the school. Diana expressed her ideas about the frequent tests and exams the school assessment structure required in our first interview:

I think the school tests is a bit too much [laugh]. Of course, my colleagues also have the same idea. Three unit tests and one examination within 6 months are quite a lot for Y7. Tests certainly have negative effects [on student learning], but for this private school, I feel that their parents are very supportive, as they hope their children could have greater academic achievements. Therefore, I feel that implementing assessment is related to the nature of the school and demands of parents. Within this context that students are certainly under the pressure of examination. (Interview 1)

Diana interpreted the parents' and school's pressure for high achievement as high achievement in tests, leading to many tests in order that students learnt to be test-takers.

Apart from using school assessments to evaluate student learning, in the second observation, Diana was observed using questionnaires to elicit students' opinions regarding their previous learning experience (see Figure 6.5).

Figure 6.5

The Questionnaire Designed by Diana to Elicit Students' Opinions About Learning Chinese
Language

Questionnaire for Yr10 Second Semester 2018

姓名：＿＿＿＿＿＿＿＿　班级：＿＿＿＿＿＿＿＿

1. Are you going to continue Chinese in Y11?　　Yes　　No

2. Do you agree on the following suggestions?

	suggestions	agree	neutral	disagree
1)	Study and revise at least half an hour each day for vocab and sentence structures.			
2)	Keep a note for topic vocabulary, buying, school, travel, etc.			
3)	Keep a diary in Chinese (twice a week to start off).			
4)	Take an active part in each class by asking questions and offering answers.			
5)	Make good use of the resources on Education Perfect, including finishing all assigned tasks there.			
6)	Make a weekly planner for activities to improve your four skills and monitor your fulfilment.			
7)	Attend a Chinese corner half an hour each week to improve speaking.			

Ideas or strategies that can help you improve on this subject:

3. Do you agree on the following requirements?

	requirements	agree	neutral	disagree
1)	Take an active part, always taking notes and concentrating in class.			
2)	Finish your Writing Book with neat, clear and accurate answers.			
3)	Finish the assigned tasks in Education Perfect in time.			
4)	Revise well before vocabulary tests and unit tests.			
5)	Value your class time and the teacher's efforts and take the subject seriously as a mature learner. (You can set a reasonable goal in Chinese subject for this half year and make good use of class time.)			

Diana explained that her intention of using a questionnaire was to elicit information about her students' learning experience in this semester. In the first section, this cohort of Year 10 students was asked which optional subjects they intended to choose for the next year. This was because Diana wanted to know their decisions and make adjustments to her teaching plans accordingly. If fewer students wished to continue learning Chinese language in the next year, Diana would design more interesting activities for students before they moved toward Year 11. On the contrary, if many students decided to continue learning Chinese language, Diana would arrange students to do more practice. She said students who decided to choose Chinese as a subject were ready for more challenges in learning.

In the second section of the questionnaire, Diana listed several learning strategies, such as keeping diaries and actively engaging in classroom activities. Diana wished to find out what the popular methods of learning were through their responses so that she could adopt these methods in future teaching practice. Diana listed her requirements of students and hoped students would commit to being actively engaged in learning in the classroom.

Interpreting Assessment Data. There was limited evidence regarding Diana's interpretation of assessment data. However, during the second lesson, Diana was observed to use a line graph to summarise students' performances in the unit test. The line graph was the distribution of all students' scores. Diana described scores of all Year 10 students as a "Year 10 phenomenon":

> This is the Year 10 phenomenon, right? What do I mean? Some just at top, and some just giving up, I think you are still here, above, you are doing a good job, cause from here above, some of you have already known that, in senior levels you don't take Chinese, but you still be working hard. (Observation 2)

Diana put the overall student performance into the specific context of a Year 10 phenomenon, in which some Year 10 students had decided to drop Chinese next year, and thus they did not put much effort in learning Chinese. She further encouraged students to keep trying to achieve better results in future exams.

Using Assessment Data. Based on different purposes of using assessment data, collected data from observations was categorised into three themes: formative uses, which included assessment for learning and assessment for informing teaching; and summative use of assessment data for recording learning during classroom activities. Each theme is discussed in the following subsections.

Formative Purpose: Assessment for Learning. During the observations, Diana was observed motivating students by giving them corrective and encouraging feedback. Like Amber, Diana frequently used "Good," and "Well done" to confirm student performance and encourage them to keep speaking in Chinese language. Furthermore, she also provided corrections when students made mistakes in speaking Chinese language. For instance,

> D: What is it?

> S: 厕所 (toilet)

> D: 厕所，is it actually a room?

S: 厕所和浴室... (toilet and bathroom)

D: 厕所和浴室 are actually the same, and 卫生间. These three rooms are actually the same, but if you want to say toilet, right?

S: Yes.

D: It is 马桶 (Observation 1)

Diana explained the differences in meanings of three Chinese words for "toilet" when she found that students had difficulties in translation.

However, Diana's assessment practices in class did not always line up with her espoused beliefs. Although Diana claimed that she was confident in using formative assessment to provide suggestions and advance students' learning in future (Interview 1), during the classroom observations, the most frequent feedback from Diana was corrections or encouragement, she did not provide suggestions for student learning during the two observations.

Formative Purpose: Informing Teaching. Before Diana finished the 1-year ITE programme, she had realised that assessment information was used to inform teaching (Interview 1). During the second observation, Diana sent out a questionnaire (See Figure 6.5) for collecting information about students' learning experience in the second observation.

She explained that the main purpose was to adjust future teaching plans according to students' responses, as "if I just follow my plans of teaching, students will resist learning Chinese." (Interview 3). However, the content in this questionnaire seemed to seek agreement from students about learning within the classroom and after class, such as "Take an active part, always taking notes and concentrating in class" and "Keep a diary in Chinese (twice a week or to start off)." Nevertheless, Diana claimed that she listed several assessment activities in the questionnaire to find students' interest in planned assessment activities in class, so that she could adopt more of those kinds to motivate students to learn Chinese language in future (Reflective Conversation 2).

Summative Purpose: Recording Learning. During both lesson observations, it was found that Diana did not always act on the assessment data she collected. Her feedback to students' questions was mostly giving direct answers or even no feedback. During the second observation, for instance, Diana asked students to read a dialogue by assigning girls to read one character and boys to read the other. When Diana noticed that the boys had difficulties

reading out their lines, she overlooked this feedback and led the boys' reading. As a result, only a few students kept up with her during the process of reading.

Another example was that Diana provided feedback by expressing personal judgement on students' performance after the unit test. During the second classroom observation, Diana spent half of lesson discussing the unit test results with students. She first showed a line graph of the score distribution and then commented on students' performances in this test:

> I think all the students in our class actually did a good job, do you understand? But some of you, I might to give you some tutorial, cause otherwise, this line is leading down, I don't know what will happen after the other term, that will be one number figure, do you understand? [Smile] It is horrible. So average is about 71%, that is average. Are you happy with your score? (Observation 2)

Diana employed a line graph to show her evaluation of the overall performances of students in the test. She attempted to prompt students to reflect on their previous learning by showing the distribution of overall scores. Nevertheless, the feedback from the unit tests was limited. She did not provide further suggestions about how students could improve learning, except to tell students that the average score was not satisfying and some students would need tutorials. After showing the line graph, Diana arranged students to correct their answers in the test:

> So when you see your mistakes in the test, you just correct your papers, right? Then I will check if you have done that. So, you have to fix all your mistakes in your test papers, for some of you, you may have more to write, you understand? (Observation 2)

Diana's instruction highlighted her requirements of students rather than providing suggestions to students about improvement. Finally, when students finished their corrections and Diana led students to work through the test content, Diana's feedback was mostly direct answers. For instance,

> D: What is it?
>
> S: I know it is 空...
>
> D: Okay, 空调。 (Observation 2)

The extract shows Diana's feedback for students was more evaluative than constructive. Compared with Michelle, who designed a lesson to discuss test results with individual students after tests, students in Diana's classes had fewer opportunities to reflect on previous performances.

Diana's Unplanned Assessment

During the two lesson observations, limited unplanned formative assessment was noticed in Diana's classroom. This was because Diana had a tight teaching schedule and she tried to finish her teaching plan in each lesson (Interview 1). Thus, students had few opportunities to interact with Diana. For instance, during the first classroom observation, Diana suggested students stop taking notes from the previous activity and move on to the next exercise.

> D: Okay, I don't think you have time to take notes of this; I will give you the handout for next period okay? Open page 10, you want to do it yourself or you want us do it together? (Observation 1)

Within the tight teaching plans, unplanned formative assessment happened when Diana noticed students had unsatisfying performances during her planned assessment activities. For instance, Diana designed a dialogue for students to practise language structures. She asked students to practise and present this dialogue in pairs. While noticing a few pairs could present the dialogue fluently, Diana realised that students still had difficulty in using in the target language structure. Thus, she adjusted the planned assessment activities instantly by providing more instructions to students:

> D: Do you want me to put the content on the board?

> Ss: Yes. (Observation 1)

Moreover, Diana was observed providing suggestions when noticing students needed support for improvement. For instance, during the second observation, Diana's questionnaire [for improving both her teaching practice and student learning in the future] provided a few strategies that students could use to learn Chinese, such as keeping diaries in Chinese and taking notes of new vocabulary. When a student asked Diana how to keep diary in Chinese language to improve learning, she replied:

> D: Write down what happened in the daytime. These are just some suggestions…maybe I think not all students can reach that level, but if you want to be like that, I think this is another way for you to improve. (Observation 2)

Diana pointed out the challenge of keeping a diary in Chinese; however, her feedback was not still clear enough for students to write diaries.

In addition, Diana also used assessment information to engage students in assessment activities. For instance,

> D: Alright, shhhh, [knocks the board] come to here. Now you are on page 3, so you give me the answer, since you are not concentrating. (Observation 2)

During the second observation, Diana recognised that students were distracted when they finished the previous activity. Therefore, she quickly moved to the next activity.

Summary of Diana's Practice of Assessment

In this section, Diana's classroom practice of planned and unplanned assessment activities has been examined and presented. The analysis of Diana's classroom observations showed a discrepancy between Diana's espoused belief of using assessment to promote student learning (Interviews 1 and 2) and her actual classroom practice, in which assessment data were overlooked or not used. This might be related to Diana's prior teaching experience of teaching in China, in which teachers were encouraged to help students to achieve in learning through intensive practice. Moreover, the factor of the school assessment requirements might have also influenced her to emphasise the summative uses of assessment in contrast to her desire to use assessment to inform learning.

Chapter Summary

Phase 2 aimed to reveal three participants' conceptions and practices of assessment as they moved towards gaining a full practising certificate. The findings in this chapter provide evidence of how they planned and conducted assessment as beginning teachers. There was evidence that the assessment practice of the three participants could be characterised within the conceptual framework, which uses three aspects to analyse and understand participants' classroom assessment practice. However, the assessment practices of the three participants varied.

At the planning stage, the three participants' classroom practice showed different characteristics. Amber realised the importance of introducing learning objectives and success criteria to students, although she was observed to keep this information for refining teaching practice rather than communicating with students. Michelle built a systematic routine to deliver teaching and assessment effectively, while Diana struggled in her understanding of the school standards when she started teaching and relied heavily on tests.

As regards eliciting and interpreting assessment data, participants had different approaches to collecting assessment information. On the one hand, all the teachers adopted various methods to elicit information regarding the extent to which students had learnt. The most frequent methods were assessment activities, questions and answers. On the other hand, when conducting summative assessment, Amber was flexible when conducting exams to ensure that students had opportunities to prepare by themselves (Observation 2); Michelle delivered unit

tests at the end of learning topics by having students complete the tasks on an online system; while Diana employed questionnaires to elicit students' feed back into test design.

With regard to using assessment data in the classroom, both formative and summative purposes of assessment were found across their planned assessment activities. Formative assessment purposes included using formative assessment data to support learning and teaching, while summative assessment purposes included recording student-learning performance, reporting achievement and moderating exam scores. Moreover, unplanned assessment was found within their teaching processes. Information from unplanned assessment activities was found to feed back into learning, informing teaching practice and engaging students in learning. However, it was found that these teachers did not always act on the assessment data they elicited or collected. Observations of Diana showed that after eliciting assessment information from students, no further action appeared to follow.

In the next chapter, the findings from Chapters 4, 5 and 6 are considered and discussed.

Chapter 7 Discussion and Conclusions

Chapter Overview

As noted, this qualitative study aimed to explore and understand the changes in Chinese-heritage language teachers' conceptions and practices of assessment longitudinally from the preservice stage to the in-service stage. The overarching question that guided this study was:

> *What are beginning Chinese-heritage language teachers' conceptions and practices of assessment at the completion of the 1-year initial teacher education programme and as they begin teaching over the first 2 years?*

Two successive studies were designed and conducted to answer this research question. Chapter 4 presented the changes made to preservice Chinese-heritage language teachers' conceptions of assessment during the 1-year ITE programme; Chapters 5 and 6 reported and illustrated the changes made by three beginning Chinese-heritage language teachers to their assessment conceptions and practices during their first 2 years of teaching.

It was found that the beginning Chinese-heritage language teachers' assessment learning did not develop smoothly from the preservice stage to the in-service stage. The findings of the Phase 1 study revealed that the participants shifted their negative attitudes toward assessment and appeared to embrace the formative assessment culture in New Zealand during the 1-year ITE programme. They expressed confidence in adopting both formative and summative assessment in classrooms to promote teaching and learning. Nevertheless, factors such as a lack of resources and support during practicum appeared to hinder their learning about and use of assessment in practice. As these beginning teachers began teaching in New Zealand secondary schools, they accumulated experience and personal practical knowledge about assessment during their own classroom practice. However, the data shows that they experienced ongoing difficulties in putting what they had learned about assessment in ITE into practice. Factors such as accountability pressure from schools and limited resources were suggested by beginning teachers as hampering their implementation of both formative and summative assessment in classrooms.

This chapter discusses the changes in Chinese-heritage language teachers' assessment conceptions and practices across their early teaching career, explores the influences affecting their assessment implementation, and proposes the potential for further research on beginning language teachers' development of assessment where cultural beliefs may also be an influence.

184

The discussion begins with the factors contributing to preservice Chinese-heritage language teachers' assessment learning in the ITE programme (Phase 1). Next, factors that influenced participants to modify their conceptions and practices of assessment in New Zealand secondary school context as beginning teachers are discussed (Phase 2). These are summarised under three subheadings, personal, school (microlevel), and cultural (macrolevel) factors. Finally, the implications and limitations of this research and suggestions for future research are addressed, before the conclusions of this study are presented.

Preservice Chinese Language Teachers' Conceptions of Assessment and Influences on Their Assessment Conceptions at the Conclusion of the ITE Programme

Based on the focus group interview investigating seven preservice Chinese-heritage language teachers' conceptions and practices of assessment at the exit of a 1-year ITE programme in the New Zealand context, Phase 1 provided evidence that teacher education can have a positive influence on preservice teachers' assessment conceptions, practices, and confidence. This finding is in line with previous research in which teachers' professional learning leads to changes in their conceptions and practices of assessment (Fenstermacher & Berliner, 1983; Richardson & Placier, 2001; Smith et al., 2013; Tam, 2013). Thus, this study provides support for DeLuca et al.'s (2013) conclusion that "pre-service teacher education has a critical role to play in promoting assessment literacy in beginning teachers and providing a foundation for teachers' continued learning about assessment throughout their careers" (p. 107).

Previous literature has found that preservice teachers often have little knowledge of assessment and low levels of confidence in using assessment at entry to teacher education (Brown, 2011; DeLuca at al., 2013; Graham, 2005; M. F. Hill et al., 2014; Volante & Fazio, 2007). In this study, the seven preservice Chinese-heritage language teachers reported their understandings, viewing assessment almost exclusively as tests or examinations at the beginning of the teacher education programme. This finding regarding preservice secondary language teachers aligns with M. F. Hill and Eyers' (2016) finding that primary preservice teachers often have feelings of fear or anxiety about assessment and find it hard to understand assessment at the beginning of their teacher education programmes, and this may be related to their prior experience and beliefs about assessment. In addition, their responses highlighted that their limited understandings and negative feelings toward assessment also appeared to have been influenced by the Chinese examination culture.

The findings of Phase 1 corroborate previous findings that the Chinese examination culture influences students and teachers' assessment conceptions and practices (Brown & Gao, 2015; Davey et al., 2007; Liu et al., 2016). Davey et al. (2007) indicate that high school teachers in the Chinese context are likely to "teach to test" due to the influence of the high-stakes Gaokao, and thus classroom assessment activities are designed to prepare students to achieve better scores in examinations. In Phase 1, preservice Chinese-heritage language teachers reflected that they used to conceive of assessment as only tests and exams when they were learning/teaching in the Chinese context, and that their prior assessment experience was mostly evaluative. They received limited feedback from summative assessment activities which were usually not oriented to their future learning.

Consistent with research that teacher education can positively shift preservice teachers' conceptions and practices of assessment (DeLuca et al., 2012; Eyers, 2014; M. F. Hill et al., 2014; M. F. Hill & Eyers, 2016; Smith et al., 2008), all the seven participants in Phase 1 reported that they had enriched and deepened their understandings about assessment, especially in formative assessment, during the 1-year teacher education programme. This finding supports previous research findings that teacher education can have positive influences on preservice teachers' conceptions of assessment, including formative assessment (DeLuca et al., 2013). An example here is that Amber illustrated the continuous influences of her ITE courses on her practices of assessment during Phase 2. As reported in Chapter 6, she said that what she had learned in the ITE programme built her teaching philosophy and set up high standards for teaching, which helped her set assessment objectives to work toward (Interview 4).

However, the changes that the preservice Chinese-heritage language teachers had made in assessment conceptions during the 1-year ITE programme were found to be positive but fragile. First, the data in Phase 2 indicated that these teachers appeared to have limited understandings about assessment and thus were likely changed by the assessment cultures in their particular secondary schools. Their limited understandings about assessment may be explained by the fact that this cohort of participants only learned about assessment within several sessions embedded within a course called Teaching Languages in the ITE programme. This finding supports previous literature on the need for more emphasis on assessment in ITE programmes and for dedicated assessment courses (DeLuca & Klinger, 2010; M. F. Hill et al., 2014). DeLuca and Klinger (2010) argue for the importance of direct assessment instruction in teacher education through their investigation of preservice teachers' conceptions and

practices of assessment in the Canadian context. Similarly, M. F. Hill et al. (2014), in New Zealand, argue that preservice teachers' assessment learning particularly benefits from the dedicated assessment course in combination with an assessment curriculum embedded in the teacher education programme and practicum. In Phase 1, there had been no specific courses or focus on assessment in the Teaching Languages course at all, but rather more general sessions on assessment in languages (in general) teaching in the 1-year ITE teacher programme. Therefore, although all the preservice teachers confirmed that formative assessment could effectively support student learning, they reflected that they were still not confident to plan and implement formative assessment in practice. Such a finding also corroborates M. F. Hill et al.'s (2014) finding that preservice teachers may need further support to encourage and motivate students to develop their assessment capability.

Second, aligned with the finding presented by Liu et al. (2016) and Tung (2019), that preservice Chinese-heritage teachers may encounter difficulties dealing with assessment practices that differ from their own culture, in this study, preservice Chinese-heritage language teachers, Diana and Ava, received less support than they expected during learning in the ITE programme. Their current experience in the New Zealand ITE programme differed from prior teaching experience in which they received more training opportunities and support toward teaching and assessment in the Chinese context. Such findings suggest that specific attention towards the differences between these two assessment systems and cultures in the Chinese language-teaching course would be helpful for preservice Chinese-heritage language teachers to gain understanding about the differences between the education and assessment systems in New Zealand and China.

Finally, half of the participants in Phase 1 of this study raised the issue that they did not receive enough support from associate teachers during their two practicum placements, and thus learning about and using assessment had been challenging. Amber mentioned that she received little feedback from the associate teacher about her assessment practice during practicum. Mia felt that her associate teachers assigned her many tasks to create teaching and assessment resources while providing little guidance about her assessment practice. Diana observed that her associate teacher misused the test before completing teaching tasks. This finding that practicum does not always support learning about assessment in ITE corroborates Eyers's (2016) finding that associate teachers are not always supportive of assessment learning during practicum. As Eyers (2016) suggests, they may not always be equipped with assessment knowledge and understandings about effective assessment practice in the New

Zealand context, even though professional development in formative assessment has been offered to in-service teachers over the last decade. McDonald (2004) supports the argument that associate teachers need to provide opportunities for preservice teachers to practise what they have been learning about in their ITE programme as well as providing constructive feedback. Thus, this thesis, in line with Eyers (2016), found that building up associate teachers' assessment capability is also critical to support preservice teachers' assessment learning.

The Phase 1 findings indicated that personal factors have an influence on the effectiveness of teacher education. This is consistent with Eyers (2014), who interviewed primary preservice teachers in New Zealand and found that preservice teachers had different degrees of metacognitive awareness and adopted a variety of self-regulatory skills to learn assessment. In the present study, participants reported different understandings of using formative and summative assessment to support student learning in practice. While most of these preservice teachers illustrated that they would use formative assessment to promote student learning, Michelle articulated her intentions to integrate both formative and summative assessment to improve learning and teaching. Her understandings align with suggestions from Bennett (2011), Black and Wiliam (2018) and Carless (2012) that both summative and formative assessment could be used to support learning within teachers' craft. The differences among individual preservice teachers' understandings of formative and summative assessment showed that the ITE programme may not have similar influences on the development of preservice teachers' assessment capabilities due to preservice teachers' personal factors. This finding also supports M. F. Hill and Eyers' (2016) claim that the changes to preservice teachers' understandings are "neither uniform nor ubiquitous" (p. 100). In fact, personal factors such as cultural background and prior experience may lead to variable outcomes of teacher education.

Beginning Chinese-Heritage Language Teachers' Conceptions and Practice of Assessment in Their First 2 Years of Teaching

In Phase 2, there was clear evidence that the beginning Chinese-heritage language teachers changed their conceptions of assessment as "learning-centered" to "teaching-centered" as they crossed boundaries from preservice stage to in-service stage. Diana claimed that the emphasis of her teaching and assessment practice was to help students achieve in exams, so that she could meet expectations from her school and parents; Michelle reflected that she was more likely to use assessment data to refine her teaching practice rather than student learning, given that she was under the pressure of retaining her job and Chinese as a subject. This

finding confirms Coombs et al.'s (2019) claim that the accountability climate in schools impacted beginning teachers' assessment approaches. The findings of their investigation regarding U.S. and Canadian preservice, beginning and in-service teachers' assessment conceptions suggest that beginning teachers change their conceptions and practice of assessment to respond to the increasing accountability climate in schools.

Furthermore, the finding that beginning secondary language teachers became more focused on summative assessment, than in their ITE programme, appears to contradict results from Brown's (2011) survey study that found secondary school teachers agreed with using assessment to improve student learning in the New Zealand context. He found that although secondary teachers conceived of assessment as formal tests, their conceptions of assessment were similar to their primary counterparts, seeing assessment as important for improvement purposes. Furthermore, he points out that both primary and secondary teachers in the New Zealand agreed more than their Hong Kong counterparts that assessment should be used to advance student learning. Based on the findings of the current study, it can be argued that certainly these beginning Chinese-heritage language teachers' assessment conceptions and practices were more similar to their Hong Kong counterparts. While they may have agreed that assessment should be used to advance student learning, in practice they centred around teachers' teaching rather than improving student learning. Their conceptions and practices of assessment were not strongly aligned with the purpose of cultivating student assessment capability, which is considered as the "heart" of teacher assessment practices in the New Zealand context (Absolum et al., 2009; Booth et al., 2014). The factors that contributed to these beginning Chinese-heritage language teachers' conceptions and practices of assessment are discussed next at the personal, school and cultural levels.

Personal Level: Conceptual Confusions Regarding Assessment

This section explores the personal factors related to the Chinese-heritage language teachers' conceptions and practices of assessment. In Phase 2 of this thesis study, the three beginning Chinese-heritage language teachers appeared to have conceptual confusions about formative assessment purposes, about the relationship between summative and formative assessment, and had limited understanding about how assessment could be both planned and interactive in nature. Carless (2005) suggests two personal factors that enable formative assessment in classrooms: teachers need to have enough knowledge about the principles and practices of formative assessment, and formative assessment knowledge needs to be aligned with teachers'

personal values and beliefs. The following subsections link with previous literature to discuss these factors.

Relationship Between Summative and Formative Assessment. One of the possible explanations for the beginning teachers' conceptions and practices of assessment is that they held conceptual confusions about the relationship between summative and formative assessment. The evidence in Phase 2 indicates that these beginning Chinese-heritage language teachers espoused that summative assessment was dominant in their decision-making regarding teaching instructions and classroom assessment. For example, Diana conceptualised formative assessment as planned assessment activities to scaffold student learning towards meeting the summative assessment requirements, and Amber conceived NCEA standards as the overarching framework to guide her design for teaching and formal unit tests. Although both of them demonstrated that exams and unit tests could be used by students to inform their learning, understanding examinations as a process of eliciting of students' proficiency at Chinese language constrained their practice of making use of examination results to construct students' future learning. Thus, their feedback on students' examination results was observed to mostly focus on student exam performance or mastery of language features. In contrast, Michelle held a clearer understanding about how she could use summative assessment results to advance student learning. She actively tried different methods to use test results to give feedback on student achievement as she began teaching and finally found one-on-one feedback was more effective to promote individual student learning. In her class, individual students were invited to have a short conversation with her after tests and exams regarding their performance, personal learning obstacles and future learning options. These findings support previous research that teachers conceive formative and summative assessment as separate forms, reducing the chances to integrate both forms in classrooms (Black & Wiliam, 2018; Deluca et al., 2012; E. Hargreaves, 2005). Taras (2007) argues that there are challenges for teachers (tutors) to understand formative and summative assessment:

> The practical and logistical problems for tutors of not having watertight definitions of the terminology and the processes of assessment are enormous. When tutors perceive that (or are led to believe that) formative assessment is something different from summative assessment, rather than a logical (often short) next step, few will jump in willingly to double their workload. (p. 370)

In Taras' view, conceptualising formative and summative assessment as two different procedures increases the perceived workload for teachers to design and implement different

assessment activities, thus teachers are reluctant to adopt formative assessment practices in classrooms. Similarly, Black and Wiliam (2018) stress that the difference between formative and summative assessment is "a distinction in the kinds of inferences being drawn from assessment outcomes" (p. 3). In their view, and in contrast to the perceptions of the beginning teachers in this study, assessment itself is not formative or summative, but how assessment outcomes will be used show the differences between summative and formative assessment functions. In other words, formative and summative purposes become evident when teachers act on assessment evidence. Furthermore, they argue that summative assessment could be used for advancing learning, which was supported by Michelle's attempts to use exam results formatively through feedback on student learning in this study. In their view, summative assessments are useful for motivating students as the owners of their own learning when the results of summative assessments are used for formative functions (Black & Wiliam, 2018).

The Spirit of Formative Assessment. In this thesis study, the three beginning teachers expressed their confidence at using formative assessment at the end of the preservice stage but were observed implementing assessment activities termed "formative" at the in-service stage. Thus, these beginning teachers conceived formative assessment as planned assessment activities and actively adopted different technologies in planning formative assessment activities. For example, Diana used questionnaires to collect students' responses about their preferred assessment activities so that she could use more of them in classrooms. Michelle constructed a routine of instructions and planned classroom assessment activities within each topic she taught. She prepared two sets of formative assessment activities, each at a different level of difficulty, to meet what she saw as students' different learning needs. The routinisation of formative assessment activities potentially diminished opportunities for her and her students to adjust teaching and learning interactively.

This finding about their understanding and use of formative assessment activities supports DeLuca et al.'s (2012) argument that teachers who believe formative assessment is activities to continuously check on progress implement formative assessment as a step-by-step routine set of activities, and thus little information from these activities may be used to improve learning. As previous literature argues, teachers often understand and practice formative assessment by following the "letter" (the strategies they have been shown or learnt about) rather than the "spirit" (Cooper & Cowie, 2009; Lee & Wiliam, 2005; Marshall & Drummond, 2006). The spirit is integrated within a clear understanding of the purposes of assessment and how these purposes might be met flexibly within a range of assessment practices, including

planned and unplanned formative assessment. Marshall and Drummond (2006) use spirit and letter to describe the difficulty that teachers encounter in transforming their conceptions of formative assessment into classroom practices:

> We use these headings—the "spirit" and "letter"—to describe the types of lessons we watched, because they have a colloquial resonance which captures the essence of the differences we observed. In common usage adhering to the spirit implies an underlying principle which does not allow a simple application of rigid technique. In contrast, sticking to the letter of a particular rule is likely to lose the underlying spirit it was intended to embody. (p. 137)

The spirit of formative assessment is consistent with Brighouse and Woods' (1999) comment about teaching. "Teachers are natural researchers, in the sense that all teaching is based on inquiry and the response of the pupils provides ready evidence as to the effectiveness of various teaching and learning approaches" (p. 42). In contrast, if teachers understand formative assessment at the letter level, they tend to control classrooms and take the responsibility to help student learning through their teaching. The possible consequence is that students lose opportunities to develop their own assessment capabilities in classroom formative assessment (Marshall & Drummond, 2006). It appears that the beginning teachers studied in Phase 2 of this thesis understood formative assessment at the letter level.

Limited Understanding About Unplanned Assessment. Another possible factor influencing the beginning teachers' operation of practices of assessment is that they were observed to lack awareness of unplanned assessment (interactive formative assessment), which is embedded in instruction or pedagogical processes. The three beginning Chinese-heritage language teachers were observed to conceptualise planned formative assessment activities as the only form of formative assessment. For example, the interview data from the three beginning Chinese-heritage language teachers showed that they conceived of planned formative assessment activity as the only form of formative assessment even though they were observed using unplanned formative interactions with students.

This finding corroborates Cowie and Bell's (1999) and Saito and Inoi's (2017) arguments that teachers are not always aware of formative assessment in their practices. In Cowie and Bell's (1999) study, formative assessment activities include both planned formative assessment and interactive formative assessment. Planned formative assessment is responsive to teachers' teaching, as this type of formative assessment is designed by the teacher and mainly used to

assess student-learning progress as specified in the curriculum. Interactive formative assessment usually happens unpredictably to mediate individual student learning. It seems more "invisible" than planned formative assessment in classrooms, as it usually happens "over a moment" (p. 112) within the processes of learning and teaching.

In this study, the beginning teachers' unplanned assessments appeared similar to Cowie and Bell's (1999) interactive formative assessment, and were found to be highly integrated with teaching and learning activities. In Phase 2, all the three beginning Chinese-heritage language teachers were observed to notice, recognise and respond to assessment information during teacher–student interactions or their observations of student-learning performances in classrooms. In other words, unplanned formative assessments were found to happen in the beginning language teachers' classroom practices. Nevertheless, none of these beginning teachers recognised their unplanned formative assessment practice as assessment. For instance, Amber considered unplanned formative assessment as a part of teaching; assessment information that she recognised during the pedagogical process was a "side benefit" of teaching and learning activities. Therefore, Amber claimed that she did not always make use of such assessment data. In other words, the teachers' own personal conceptions and beliefs about assessment could have limited their perceptions of assessment, and thus restricting development of their practices.

Cowie and Bell (1999) also suggest that both experienced and beginning teachers may have difficulties in using interactive formative assessment when they are under pressure in a new classroom environment, adopting a new curriculum or under the pressure of summative assessment requirements. That is to say, the beginning teachers' unplanned formative assessment practice is influenced by both personal factors and school-level factors. Lee and Wiliam (2005) argue that the pressure from external assessment requirements, and support from teacher development and school communities to allow teachers to sustain reflections about formative assessment, also influence their classroom practices. Al Sawafi (2014) links teachers' formative assessment practices to the innovation of assessment policies and culture and points out the influences from assessment policies and culture also influence teachers' unplanned formative assessment practices. Based on this understanding, factors from micro- and macrolevels are discussed in the following sections.

Microlevel Factor: Building Up Assessment Learning Community at a School Level

Teachers' professional learning in assessment closely relates to their school assessment environment, including the internal and external school support (Carless, 2005; Cooper & Cowie, 2010; M. F. Hill, 2011). The changes of conceptions and practice of assessment of the three beginning Chinese-heritage language teachers during their first 2 years of teaching confirm M. F. Hill's (2011) argument that school-level factors play a critical role in linking personal and macrolevel factors of teachers' assessment learning. Teachers' development of assessment knowledge and practice is related to their knowledge of students, curriculum content, and the specific school context (M. F. Hill & Jiang 2018; Shepard, 2000). In this thesis study, school-level factors were also found to be critical in supporting and causing the changes that participants had made in conceptions and practices of assessment during their first 2 years of teaching. Both Michelle and Diana were pushed by school-context circumstances to focus more on classroom management and summative assessment, and also short assessment activities, leading them away from formative practice, especially interactive formative assessment.

New Zealand has been engaged in implementing a shift in practices of assessment, from a traditional focus on examination results to embedded assessment practice in and across schools (Absolum et al., 2009; Cooper & Cowie, 2010; M. F. Hill, 2011). However, previous research found that compared with the primary sector, implementing formative and embedded assessment practice in secondary schools is more challenging (M. F. Hill, 2011; Timperley et al., 2007; Yates & Johnston, 2018). M. F. Hill (2011) studied three school cases where formative assessment practices were successfully integrated into teacher practice and school culture, and found that factors related to the principal's assessment understanding, management involvement, external expert support and school culture are critical in supporting teachers to learn and use assessment effectively in practice.

The data in the Phase 2 of this thesis demonstrated that the beginning Chinese-heritage language teachers received insufficient development for understanding and using formative and embedded assessment as they began teaching in New Zealand secondary schools. Participants' secondary schools appeared to put the emphasis on student achievements in qualification assessments at senior levels, as better student achievement can make schools "look better" (Crooks, 2011, p. 76). Therefore, beginning Chinese-heritage language teachers' conceptions and practices of assessment were influenced by such school values. In Phase 2, Diana and Michelle clearly felt the pressure from establishing teacher accountability in private

schools, while Amber felt less stressed in the integrated school. Additionally, there was little support for these Chinese language teachers to learn how to integrate formative and summative assessment strategies in practice within these schools. All the beginning Chinese-heritage language teachers struggled to learn about and to use assessment within tight teaching schedules. They learned assessment mainly from online resources or communications with colleagues. Nevertheless, their communications with other colleagues were found to be infrequent due to their cultural differences and language barriers.

Consistent with the findings of this thesis, Nolen et al. (2011) argue that beginning teachers face challenges regarding appropriation, negotiation and recontextualisation of assessment tools and practices when they move from university to schools, as well as the change that they face in their identities from students to novice teachers. Secondary schools need to recognise that beginning teachers may still need much support to develop their assessment capabilities and provide more opportunities for them to learn about assessment continuously. The following subsections discuss the support and challenges that the beginning Chinese-heritage language teachers experienced in practising assessment in the New Zealand secondary school context.

Misalignment of Assessment Priorities in Teacher Education and Secondary Schools. In this study, participants in both phases of the investigation illustrated their concerns regarding teacher accountability in secondary schools. In the Phase 1 study, one of the participants, Mia, observed her mentor's assessment practices during practicum and indicated that she might not be able to use formative assessment in her own practices as the school has an emphasis on examination results. This concern became real when the teachers began teaching after graduation. In Phase 2, the three beginning teachers demonstrated that fulfilling accountability demands was their priority, in spite of their strong motivation to use formative assessment in classrooms. Both Diana and Michelle mentioned that they were under pressure to assist students to achieve well in exams in response to expectations from schools, parents, mentors and students. This finding is consistent with M. F. Hill's (2000) finding in New Zealand primary schools that teachers' assessment practices are influenced by the competing discourses of school accountability and professional development in which formative assessment was a priority. In this present study, the data extends her findings to the New Zealand secondary school context, showing that these beginning secondary teachers experienced dissonance between formative assessment promoted in teacher education and the surveillance of accountability in these secondary schools.

Previous literature has shown that schools tend to work to raise standards of student achievement to meet increased school accountability requirements, and this may lead to the result that teachers shift their practices towards traditional forms of summative assessment (Deluca et al., 2012; M. F. Hill, 2000, 2016). Brown (2011) reminds us that emphasising school accountability over improvement is counterproductive in teachers' conceptions of assessment. Fullan (2005) and M. F. Hill (2011) also note that promoting formative assessment practices requires alignment of the education system, school sectors and personal stance. In Phase 2 of this study, Michelle and Diana seemed to change their priorities from using assessment to support student learning to using assessment to meet increased accountability demands from schools. As a result, assessment information was used mostly for teaching, deviating from the formative assessment purpose of supporting learning, the beginning teachers were pressured to change their assessment practices shaped in the ITE programme to fit into school assessment cultures (e.g., Absolum et al., 2007; Brown, 2002; Cowie & Bell, 1999; Cooper & Cowie, 2010; M. F. Hill, 2000).

Limited Time and Resources. Another factor that influenced the beginning teachers' implementation of assessment in these classrooms was limited time and resources. All the beginning teachers demonstrated that they were overwhelmed by the need to create assessment activities and design valid tasks and test papers throughout their first 2 years of teaching. This finding aligns with previous literature that the lack of time and resources constrains teachers from implementing formative assessment in classrooms (Carless, 2005; A. Hargreaves et al., 2002; M. F. Hill, 2011; Inbar-Lourie & Donitsa-Schmidt, 2009; Tung, 2019). Inbar-Lourie and Donitsa-Schmidt (2009) found the lack of time and lack of resources were predictors of teachers not implementing formative assessment in classrooms. The three beginning teachers in the current study reported that although they had strong motivation to create assessment activities to engage students in classroom learning, creating such activities doubled their workloads. The tension between their intentions and beliefs and reality confirms that beginning Chinese-heritage teachers need more resources and time to design and implement formative assessment.

However, as Deluca et al. (2012) argue, implementing formative assessment in classrooms does not necessarily require resources and technologies, because the focus of formative assessment can also be unplanned interactions between teachers and students. The three beginning teachers reported formative assessment practices were confined to creating various formative assessment activities, confirming the evidence that the beginning teachers still

lacked awareness and understanding of the spirit of formative assessment, and neither recognised nor acknowledged the potential of unplanned assessment.

Furthermore, in New Zealand, *learning languages* was established as a learning area in the national curriculum in 2006, and the development of language programmes in secondary schools is still in progress (East, 2008; East & Ward, 2016). Compared with mainstream subject teachers, these beginning Chinese-heritage language teachers were challenged to create both teaching and assessment resources without reference to previous well-established resources. Michelle commented that she took lots of time to prepare teaching and assessment resources for Chinese language classes. In contrast, she felt it was much easier to prepare for social studies classes, given that resources for the social studies subject have been established and used in her school for many years.

Principal Support of the Change in Assessment Culture. In this study, different findings in three participants' schools illustrated the critical roles of secondary school principals in building a school culture of formative assessment. For example, Michelle reported that her school changed the strategy of examination feedback based on teachers' advice. As she explained, the school used to ask teachers to provide written feedback on student examination results and thus increased the workload for teachers after each test. Teachers may also miss the right timing to communicate with individual students about their performances in tests. Thus, the principal in her school had changed the strategy and offered more liberty for teachers to choose what they thought more effective in giving feedback. Therefore, Michelle adopted using test results to give feedback to inform student learning and guided students to make use of test results in future learning. Similarly, Amber reflected that she received support from the principal to design and implement assessment based on her judgement.

Findings from this thesis corroborate M. F. Hill's (2011) argument that embedding and sustaining formative assessment practices across secondary schools requires principals to have deep understanding of assessment, and formative assessment in particular, and have full involvement in promoting formative assessment practices across schools. Chandan (2018) suggests that formative assessment can be most effective when teachers take the initiative and responsibility in assessment. Jones and Webb (2006) further argue that the factors contributing to sustaining formative assessment practices and supporting changes that teachers made regarding formative assessment came from the school leadership, including distributed leadership. In Michelle and Amber's schools, the principals supported teachers to

take initiatives in their profession. For example, based on the teachers' discussion, Michelle's school changed the requirement for teachers to provide written feedback after each exam. Therefore, Michelle had more time to provide feedback to individual students.

Pressure from Tests and Exams. In Phase 2 of this study, the influence from tests and examinations on Diana's assessment practice significantly increased when Year 10 students were choosing subjects for the next year. Diana reflected that many students decided to drop Chinese as a subject in Year 11 in order to choose a subject in which they could show better achievement in the qualification assessments. Thus, she started to plan more language input and increased the challenges of learning to ensure students achieved well in high-stakes assessments.

Studies have shown that New Zealand secondary school teachers' conceptions of assessment are influenced by high-stakes examinations (Edwards, 2017; Yates & Johnston, 2018). Yates and Johnston (2018) argue that secondary school teachers are more likely to link exam results with student accountability and thus put emphasis on improving student exam results. They analysed 115 New Zealand in-service secondary teachers' responses regarding assessment and found that there is weak relationship between formative assessment and NCEA school-based exams. As they argue, teachers are under the pressure of high-stakes NCEA exams, as their teaching quality and school accountabilities are also evaluated based on these exams. As a result, teachers' feedback on summative assessment results may be oriented to improve student future exam results rather than improve learning and thus lose sight of the efficacy of formative assessment. In Phase 2, Diana was observed to provide evaluative feedback on students' performances after a unit test. She provided correct answers to the unit test and asked students who did not perform well in the unit test to attend her tutorials in future. Thus, assessment information that she gained from the unit test was not used to guide student future learning. Diana's assessment experience was also aligned with Edwards's (2017) findings, in which beginning teachers may place their emphasis on using exam results to adjust teaching so that they can better prepare students for future examinations.

Also, Diana's school offered students learning programmes in both IB and CAIE. Well-established IB and CAIE assessment systems cause potential peer competition worldwide, and this may be helpful to explain Diana's adoption of intensive teaching and test strategies to prepare students for peer competitions in such high-stakes examinations.

Another possible factor that may contribute to the participants' responses to high-stakes assessments is that their school communities are different. While Amber worked in an integrated school, which she found to be relatively low stress, Diana and Michelle worked in upper-income private school communities. In their communities, they believed that student achievement received more attention from the school, parents and students themselves. Therefore, they felt very responsible to assist students to achieve high scores in exams. M. F. Hill (2000) argues that increased managerial accountability in schools is due to the marketisation of education. As a result, Michelle and Diana changed their teaching and assessment focus to respond to what they perceived to be the expectations of student achievement from parents, the school and even the students themselves.

Support from Internal and External Facilitators. In Phase 2, the findings that the beginning Chinese-heritage language teachers in this thesis struggled to learn about assessment and create teaching and assessment resources suggests that there is a link between the beginning Chinese-heritage language teachers' continuous learning needs regarding assessment and insufficient support within and across schools. This finding corroborates previous findings that practising teachers need opportunities to reflect upon classroom practices and make connections to theory (Booth et al., 2014; M. F. Hill, 2011; Torrance & Pryor, 2001). Booth et al. (2014) point out that teachers need to actively engage in professional learning so that they can "change or enhance their teaching, learning and assessment practices" (p. 148). In the present study, while the beginning Chinese-heritage language teachers realised the importance of continuing to learn about assessment, crowded teaching schedules and limited opportunities of learning assessment constrained their reflections about assessment practices. For instance, in Phase 2, both Diana and Michelle reflected that they had only two ways to learn assessment: online resources and interactions with other subject teachers. However, they claimed, searching suitable online resources took lots of their time. In addition, Diana appeared to experience more frustration in communicating with other subject teachers in the school due to the language barrier and cultural differences between her and other colleagues. M. F. Hill (2011) suggests that customisation of professional development in each secondary school could empower teachers to use effective formative assessment through collaborative inquiry and communications with teachers who are more assessment literate. However, in this study, only Amber reflected that her school held workshops for teachers to share teaching experiences; both Michelle and Diana had no such learning opportunities within their schools.

Furthermore, the findings of the current study suggest that the number of workshops and conferences provided by the teacher induction programme were found to be insufficient to support these teachers to solve their assessment challenges in the classroom. In Phase 2, Michelle was constrained by a tight teaching schedule and thus missed the opportunity to attend such workshops. Diana and Amber shared a similar view that cross-school communications were important for the beginning Chinese language teachers to share resources and learn from experienced teachers. Diana compared her teaching experience in the New Zealand secondary school context and prior teaching experience in the Chinese context and pointed out that an organisation such as teaching and research groups in each city of China provided frequent support to teachers across schools. Regular workshops and conferences, observations of other teachers' classrooms and connections with researchers allowed teachers like her to develop professional capability continuously but she did not experience that sort of support in her first 2 years of teaching in New Zealand. While Cooper and Cowie (2010) stress that external support is important to maintain teachers' focus on formative assessment by developing teachers' assessment practices, the findings in the present study suggest that sufficient external support based on beginning Chinese-heritage language teachers' learning needs could effectively assist them to learn assessment continuously.

Macrolevel Factors: Influences from both the New Zealand Context and Confucian-Heritage Cultures

The literature speaks of macrolevel factors as a contributor to changes in language teachers' conceptions and practice of teaching and assessment (Carless, 2005; East, 2006a; M. F. Hill, 2011; Ward & East, 2016). Carless (2005) lists a reform climate, social teaching, the learning and assessment culture, the impact of government or quasi-governmental agencies, and the role of high-stakes tests as factors which could contribute to language teachers' conceptions and practice. In this thesis study, both New Zealand and Chinese teaching and assessment contexts were found to have impacted the beginning Chinese-heritage language teachers' assessment conceptions and practice. These influences are discussed next with reference to previous literature.

Influences from the New Zealand Language-Teaching Context. In this study, all the preservice and beginning Chinese language teachers who participated explained how they experienced limited resources impacting upon their assessment practices. As noted above, in Phase 1, the preservice teachers reflected that they had limited resources to draw on to learn about assessment as well as little support from associate teachers during practicum, and thus

held regular meetings within the cohort to support each other. In Phase 2, both Michelle and Diana demonstrated their assessment practice was constrained by the pressure of retaining enough students in their Chinese language classes. Furthermore, Michelle reflected that the abstract assessment standards in the curriculum documents, and in her school, for Chinese increased the challenge to create teaching and assessment materials. Similarly, Amber was frustrated by the fact that she needed to both decide the curriculum and design assessment for Chinese with less than satisfying support from the education system. She commented that the development of Chinese as a subject was still in progress in the New Zealand context, which increased the challenges for the beginning teachers to teach and assess effectively.

The findings of this study corroborate previous studies that have found languages subjects (other than English) lack support at the national policy level (Ashton, 2018; East, 2006, 2018; Ward & East, 2016). Since learning languages was established as a new learning area in the revised *New Zealand Curriculum* in 2006, students from Years 7–10 can choose to learn additional languages (languages other than English) as 'taster options.' East (2008) compared the percentage of students taking up second and additional languages from Years 9–13 in both English and New Zealand contexts, which suggest that students in New Zealand schools have a low uptake of languages. He argues that the lack of compulsion to learn second/foreign languages in schools has hindered the development of language subjects. In fact, from Year 9, some schools only offer one or two terms of other language learning, rather than a whole academic year, under such policy orientation.

Furthermore, in Phase 2, all the beginning Chinese-heritage language teachers were observed adopting various teaching and assessment methods and techniques in language classrooms. They explained that they prioritised using interesting activities to attract students to continue learning languages. Despite the effort they made to keep students learning Chinese language, Diana commented that there were always only one or two students in their Year 13 classes, as few students have the confidence to learn additional languages. East (2000) and May (2002) argue that the belief that considering English as the only important language has hindered the development of second languages at the social-cultural level. Therefore, influenced by the factors of the low uptake of students learning second languages in schools, and the lack of a culture of valuing second/other language teaching at school level, second/other language teachers are "frequently facing demotivated students, low enrolments and considerable attrition" (East, 2012, p. 128).

Influence from the Confucian-Heritage Cultures. Beyond the discussions about the changes of beginning Chinese-heritage language teachers' conceptions and practices of assessment within their personal stance and school level (microlevel), influences from the CHCs on beginning Chinese-heritage language teachers' assessment conceptions and practices are discussed in this section. In the New Zealand context, cultivating students' assessment capabilities is considered as "the heart of assessment" (Booth et al., 2014, p. 137) in teachers' assessment practices. Students' involvement in assessment is a critical way to develop their capability in assessment and further supports them to become lifelong learners (Absolum et al., 2009; Booth et al., 2014). However, in Phase 2 of this study, the three beginning teachers demonstrated that their focus was on using formative assessment to refine their own teaching (such as Diana's practices of recording assessment information) rather than to empower students to use assessment to improve their own learning. The observation data indicated that feedback upon student performances in assessment activities was mostly corrective and encouraging comments; comments were rarely used to encourage and enable students to assess their own learning. This finding aligns with Zhu and Edwards's (2019) finding that CHC teachers are likely to assess students through low-level questions about their mastery of knowledge rather than supporting them to think for themselves through more open or exploratory questioning. They explained that CHC teachers lack motivation to change their methods of inquiry due to the pressure from the examination culture in China. The beginning Chinese-heritage language teachers in this study were observed to frequently use questions to check students' mastery of target language features, and their comments on students' responses were mostly corrective. In this way, they could ensure that students had learnt the target language features.

Furthermore, during the assessment activities, the beginning Chinese-heritage language teachers still took a dominant place in assessment activities, and thus students had limited opportunities to engage in classroom assessment decision-making processes. The discrepancy between the beginning teachers' conceptions of formative assessment and advocated assessment practices in the New Zealand context indicated that the beginning teachers had limited understanding of the real purpose of cultivating student assessment capabilities. The difficulties that the beginning teachers encountered in understanding and implementing advocated assessment practices may be explained by the cultural differences between the teacher's role in CHCs and in the New Zealand assessment culture.

Teachers' Roles in Assessment in CHCs. CHCs value teachers' roles in hierarchical relationships with students and school communities (Corcoran, 2014; Cortazzi & Jin; 1996; Davison, 2004; Perez & Shin, 2016; Sun, 2010). First, teachers in CHCs are more likely to control student learning in classrooms, as they are highly responsible for student-learning outcomes and are expected to deliver knowledge and skills. Cortazzi and Jin (1996) summarise the different characteristics of learning between the CHC context and Western countries. They indicate that in the cultures of hierarchical relationships, students show more agreement to teachers' decisions; teacher–student communications are more likely to be formal. Perez and Shin (2016) examined the learning style of students in the Chinese university context and found that students are not likely to interact with teachers in such relationships. In Phase 2, both Amber and Diana expressed that they attempted to control student learning in classrooms at the beginning of teaching. They believed that they took high responsibility for students' learning outcomes, and thus attempted to control classes so that they could cover more teaching content.

Second, influenced by CHCs in which sustaining harmony in communities and diligence in work is important, teachers are more likely to change practices to meet the value or expectation of school communities. The hybrid of these two cultures is noted by literature (Fong, 2007; Kim et al., 2017) as the integration of collectivism (obedience and social connections) and individualism (independence and excellence). Teachers working within the hybrid culture in China are likely to adopt school requirements of assessment mechanically in the classroom and help students to achieve in assessment regardless of whether students may have different learning needs.

In the current study, Diana's statement of changing assessment practices to meet school assessment requirements may demonstrate her intentions to keep harmony with the school community by working hard to meet expectations of the school and parents, as her prior experience of teaching in CHCs cultures had influenced her to do so. Michelle and Amber built up the routines of formative assessment activities in classrooms, showing that they were motivated to take responsibility to advance student learning from their own perspectives. Interestingly, Michelle claimed that she adopted various formative assessment activities to keep students motivated in learning Chinese, due to her perception that there is a stereotypical view of Chinese-heritage teachers as only able to use mechanical drills to teach. Michelle's statement showed evidence that as a beginning Chinese-heritage teacher, she attempted to change practices to meet expectations from students, teachers, schools.

Influence from Examination Culture. Another possible influence is that the long-existing examination traditions in CHCs may still have an impact on the beginning Chinese-heritage language teachers' conceptions and practices of assessment. In the present study, both Amber and Diana seemed to be less confident in designing summative assessment activities. They preferred to adopt or adapt well-established test papers from experts, as they had experienced in the Chinese context. Furthermore, Diana reflected that her responsibility was to help students achieve higher scores in exams. Their assessment experience showed the beginning Chinese-heritage language teachers were likely to be influenced by their prior experience in the examination culture of the Chinese context.

Formative assessment has been recognised by policymakers and has been encouraged for use in classrooms in the Chinese context for many years; nevertheless, the literature suggests that the examination culture still prevails in influencing teachers' classroom practices (Chen, 2015; Chen & Brown, 2016; Gu, 2014; Poole, 2016). Chen and Brown (2016) argue that in the CHCs, high-stakes tests and examinations are considered as a force for improvement of students' learning. Thus, Chinese teachers are likely to conceive summative assessment as a way to establish both student accountability and learning improvement. Aligned with Chen and Brown's (2016) argument, both Amber and Diana in Phase 2 adopted frequent summative assessment activities to assess and supervise student learning. Diana conceived summative assessment activities as a way to help students take responsibility for their own learning. Furthermore, this finding also supports Gu's (2014) argument in which summative assessment is not inherently bad. The challenge of assessment innovations for teachers and schools is to find the balancing points to use different types of assessment appropriately to support student learning and encourage students to take responsibility for their own assessment. In Phase 2, Michelle was observed to communicate with individual students about their exam results and learning difficulties, showing her attempts to use both formative and summative assessment to support student learning. However, none of these beginning teachers were observed supporting their students to take agency for their own assessment.

Influences from the Behaviouristic Approach to Teaching and Learning. The beginning teachers' development of their assessment conceptions and practices may have been influenced by the differences between their CHCs and the New Zealand educational culture. For example, the beginning Chinese-heritage teachers felt insecure without well-established exam papers and textbooks, and ready-to-use resources for teaching as they began to teach in New Zealand secondary schools. The difficulties that the beginning Chinese-heritage

language teachers encountered may be due to their prior experience of learning and teaching structured by a teaching and assessment approach underpinned by behaviourist principles in the Chinese context, and thus they appeared to lack understanding and skills to take the initiative in creating resources for teaching, learning and assessment to meet student learning needs.

In the Chinese context, the didactic approach based upon behaviouristic theories of learning is widely applied in teaching and assessment (Carless, 2005; Corcoran, 2014; Zheng, 2013). Cortazzi and Jin (1996) assert that the method of teacher-centred knowledge transmission is commonly found in the Chinese-heritage culture. Tan and Chua (2015) claim that under the education reform in the Chinese context, teachers tend to return to prior cultural emphasis, in which "teacher-directed teaching and knowledge transmission tend to rely on high-stakes examinations to test learning outcomes, and adopt uniform standards to appraise students" (p. 695). Similarly, Lam (2016) commented that language teachers experience frustrations to cultivate student learning in the CHC context where "language teaching (for example, teaching of writing) equates to examination preparation" (p. 766). In the current study, Diana struggled to cover teaching content as she began to teach. As she reflected, she did not know that teachers could adjust teaching and assessment based on student-learning progress, and thus she attempted to control her classes and teach more to meet teaching plans at the beginning of the semester. This finding supports Sun's (2010) argument, in which teachers from CHCs may encounter difficulties in classroom teaching when they hold the belief that "a 'good' teacher is always expected to present content in an orderly and smooth fashion" (p. 227).

In the New Zealand context, influences from constructivist beliefs on teaching and assessment are often documented in studies regarding science education. Garbett (2011) indicates that science teaching relies on a personal constructivist view in the New Zealand context. Student learning is highly related to teachers' capabilities and confidence:

> Having the confidence and ability to direct students' learning in a constructivist way depends on the teacher's own sense of confidence to manage the learning environment safely and competently and to be able to deal with observations and questions from students which are unexpected. (p. 39)

Some time ago, Bell and Gilbert (1996) asserted that teachers who adopt the constructivist approach to teaching are likely to take into account students' existing ideas, creating

opportunities to engage students in meaningful dialogues and then make use of these dialogues to interact with student thinking. They indicate that the constructivist view of learning in the New Zealand context is centred around students' personal stances. Students take initiative to construct, link and examine input with their existing knowledge. The teacher's role is to find students' prior knowledge, present and explain subject-related ideas, respond to students' ideas, encourage students to reflect on learning, and assess their changes in subject-related ideas. In other words, teachers adopting a constructivist approach are likely to support students to take initiative in their learning. In Phase 2, the beginning Chinese-heritage language teachers' assessment conceptions and practices highlighted that they used assessment information to improve teaching rather than support student learning. Both Michelle and Amber clarified that their emphasis was on using assessment to reflect on their teaching practices, showing that the beginning Chinese-heritage language teachers were not likely to adopt a constructivist approach in their teaching and assessment practices.

The Interactions Between Prior and Current Beliefs. Although prior experience of learning and teaching may influence understanding and hinder the implementation of formative assessment in the New Zealand context, the beginning Chinese-heritage language teachers were observed to actively reconcile the dichotomy between their prior and current beliefs. For example, the three beginning teachers kept trying different strategies of formative assessment activities to engage students in classroom learning and spent time on preparing students when examination dates came around. Amber demonstrated that she tried to integrate both summative and formative approaches in classroom teaching and gave students opportunities to learn independently to motivate students to take responsibility for their own learning.

Zheng (2013) suggests that teachers' development is non-linear. They may adopt a compromise to resolve the conflict of beliefs. Zheng examined the relationship between six Chinese EFL teachers' beliefs regarding teaching, learning and assessment and classroom practices in the Chinese context. She found that EFL teachers held core beliefs of adopting behaviouristic approaches to help students achieve well in exams and also the peripheral belief of improving students' language competence through a formative assessment approach, and thus adopted an eclectic approach of two kinds of beliefs in their classroom practices to achieve different objectives. Zheng argues that due to the dynamic and complex interactions between teachers' core and peripheral beliefs, adopting an eclectic approach allowed these teachers to choose teaching (and assessment) principles and approaches based on specific teaching (and assessment) contexts, reducing the influence of the high-stakes assessment on

their classroom practices in CHCs. Although Zheng's (2015) study is about Chinese EFL teachers' beliefs about teaching, learning and assessment in the Chinese context, the findings of this thesis study also suggest that the beginning Chinese-heritage language teachers attempted to actively reconcile the dichotomy between their CHC experience and the New Zealand educational culture they had been learning within, showing the interactions between their prior and current beliefs. Thus, it can be argued that an eclectic approach was also adopted by the beginning Chinese-heritage language teachers to explore suitable ways for themselves to teach in the New Zealand context. As Amber claimed, she had been looking for ways to combine different approaches of learning, teaching, and assessment from both New Zealand and China contexts to support student learning.

Furthermore, as Tung (2019) and Tan (2017) argue, CHCs are "open traditions," which are "diverse, fluid and evolving, offering an educational paradigm that is all-rounded, ethical, universal and ultimately enduring" (p. 11). Tung (2019) and Saito and Inoi (2017) assert that teachers from CHCs are not afraid of changes in the assessment paradigm. In contrast, they attempt to try formative assessment in classrooms. Their attempts in reconciling the dissonance between formative assessment and an exam-centred tradition may imply the open characteristics of CHCs and hold the promise of more change to come as they become more experienced teachers in the New Zealand context.

The Conceptual Framework of the Current Study

In the present study, a described conceptual framework of assessment practice was employed to identify and analyse the beginning Chinese-heritage language teachers' classroom assessment practices in Phase 2. Three aspects of assessment (processes, purposes and forms) were described from previous studies: Assessment purposes include both formative and summative purposes; assessment forms consist of planned and unplanned forms. With regard to assessment process, four steps of planning, eliciting, interpreting and using assessment data were described for planned assessment processes, while three steps of noticing, recognising and responding were described for unplanned/interactive formative assessment processes.

There are three significant features of the conceptual framework. First, the conceptual framework constructed and used to guide this study integrated assessment forms, processes and purposes, showing the complex interactional nature of teachers' classroom assessment practice. While previous studies have illustrated that it is difficult to separate formative and summative purposes of classroom assessment activities (Black & Wiliam, 2018; Taras, 2017),

few frameworks in the existing literature integrate formative and summative assessment purposes and corresponding processes that comprise classroom assessment activities. As a result, observing and understanding teachers' classroom assessment practices remains challenging. Black and Wiliam (2018) proposed a framework of assessment in relation to pedagogy (see Figure 2.2), in which both formative and summative assessment purposes of assessment are integrated into teachers' implementation of assessment. While their framework provides a broad view of the process of classroom assessment activities as the integration of assessment within pedagogy, it has less to say about how to use assessment information to serve formative and summative purposes. In this thesis study, the beginning Chinese language teachers' conceptions and practices were observed and described using three aspects (assessment forms, processes and purposes) at each stage as they moved toward full registered teachers. The conceptual framework of this study was helpful to explore the beginning Chinese-heritage language teachers' professional growth at each phase, in depth, and to describe differences in their conceptions and practices between stages during the investigation.

Furthermore, the conceptual framework constructed from the literature for this study extends assessment frameworks from previous studies by capturing classroom assessment activities through their planned and unplanned forms. Using planned and unplanned forms of assessment to observe and identify the beginning teachers' actual classroom activities was found useful in this study, especially when classroom activities were embedded in teaching or learning activities. As the findings demonstrated, although these teachers were aware of, designed and used planned formative assessment activities, they were less aware of their unplanned assessment activities and rarely mentioned how students themselves could and should be involved as partners in assessment of their own learning and achievement.

In addition, the findings of the current study suggest that Cowie and Bell's (1999) framework of interactive formative assessment can be used to observe and understand language teachers' unplanned formative assessment practices. During classroom observations, the beginning Chinese-heritage language teachers' unplanned formative assessment practices were found including steps of noticing, recognising and responding, providing empirical evidence to employ the framework of interactive formative assessment to explore processes of language teachers' assessment practices.

This framework could be used beyond its implementation as a research tool to observe teachers' assessment practices. It attempts to guide teachers to reflect on their assessment to gain understanding regarding different aspects of assessment, and further develop their assessment capabilities.

Integrating Methods of Interview and Observation in this Longitudinal Study

One significant feature of this longitudinal study was that the investigation followed the beginning teachers' professional growth with regard to assessment from the preservice stage through their early career of teaching. DeLuca, Willis, et al. (2019) found that beginning teachers' assessment conceptions changed significantly from the preservice stage to their early teaching career; nevertheless, there seem to be limited studies focusing on beginning teachers' professional development of assessment in their early career of teaching. Furthermore, previous studies mainly investigated teachers' assessment experience in teacher education or the beginning stage as they start teaching in schools. For example, Eyers (2014) explores changes of preservice primary teachers' assessment conceptions during coursework and practicum, and Edwards (2017) investigates beginning teachers' development of summative assessment literacy from teacher education into teaching by employing the approach of interviews and observations in New Zealand secondary schools. However, Edwards's study (2017) focuses on the first 6 months of beginning teachers' summative assessment practices in science classes. This 2-year investigation explored the beginning Chinese-heritage teachers' conceptions and practices of assessment over a longer timeframe as they moved towards a full practising certificate. Adopting approaches of observations and interviews over an extended period was helpful to trace changes that beginning teachers made in classroom assessment and to further understand factors that prompted them to make these changes over time.

Implications: Building up a Professional Assessment Learning Community

The findings of the present study confirm Farrell's (2012) note that novice teachers usually face the challenge of having limited contact with other professionals after teacher education and often experience insufficient support from schools. The beginning Chinese-heritage language teachers' struggles in practising assessment were in relation to the fact that they did not receive enough support from associate teachers during the ITE programme. Furthermore, as they began to teach, the beginning Chinese language teachers faced the challenge of learning assessment continuously within tight time schedules of teaching, school emphasis on

student achievement in exams, limited teaching and assessment resources in the Chinese subject, and limited opportunities to learn assessment within and across schools. Such findings support Tierney's (2006) argument that changing teachers' assessment capability cannot be accomplished just through professional development; it involves a shift in conceptions and practices for all stakeholders. In addition, these Chinese-heritage beginning teachers were caught between the competing cultural factors of their experience in two culturally different education systems. The very fact that they were teaching Chinese language, with all the difficulties this entailed, also raised the cultural issues related to learning another (not dominant) language. Based on the findings of this study, there are important implications for teacher education programmes for language teachers, secondary schools, and beginning Chinese-heritage language teachers themselves and these are presented in the following subsections.

Implications for Language Teacher Education Programmes

Several implications for improving preservice Chinese-heritage language teachers' assessment capabilities during the 1-year ITE programme are suggested from this study. First, the findings of this study showed that beginning Chinese-heritage language teachers seemed to have limited understanding and skills of formative assessment and thus they were more likely to shift their conceptions and practices from learner-centred to teacher-centred as they began teaching in New Zealand secondary schools. This finding supports previous research that preservice teachers need more opportunities to learn systematically about assessment in ITE (for example, DeLuca & Klinger, 2010; M. F. Hill et al., 2014), suggesting teacher education programmes could provide a dedicated assessment course for preservice teachers.

Second, the beginning Chinese-heritage language teachers were found to lack awareness that unplanned formative assessment can and should be embedded within the teaching and learning process. Their confusion about this type of formative assessment may lead them to overlook information about their students' learning and further miss the chance to adjust teaching and learning in the moment. From this perspective, a systematic review of formative assessment, including both planned and unplanned assessment, could deepen preservice Chinese-heritage language teachers' understandings of formative assessment during teacher education programmes. Teacher educators could help preservice teachers articulate their unplanned formative assessment experience and opportunities when reflecting upon practica and prior experience of assessment, deepening their awareness of making use of unplanned formative assessment data in practice.

Third, the Chinese-heritage language teachers may appreciate teacher education providing clarifications of the advocated assessment practices in the New Zealand context by comparing different assessment approaches used in China and New Zealand education systems. In this process, Chinese-heritage language teachers may need more opportunities to engage cognitively with, and reflect upon, their prior experience and understanding of teaching and assessment in CHCs. Through reflections about their prior experience, Chinese-heritage teachers could better understand the differences between the two assessment cultures and thus deepen their understandings of advocated assessment practice in the New Zealand context.

Fourth, preservice and beginning Chinese-heritage language teachers reported that they received more learning opportunities within and across schools when they were teaching in the Chinese context. Such findings suggest more research and development teams are needed across schools, as in China, to support beginning Chinese language teachers as they lack resources, assessment ideas and collegial support for teaching this subject.

Finally, the beginning Chinese-heritage language teachers may still need support to clarify the relationship between formative and summative assessment. Instead of stressing the differences between these two purposes, it is important to show preservice teachers how to use both formative and summative assessment activities to support student learning. Furthermore, the findings of this study aligned with DeLuca and Klinger's (2010) finding, suggesting that Chinese-heritage language teachers need direct assessment instruction from teacher educators and associate teachers in schools about designing, implementing, and using both formative and summative assessment activities in classrooms.

Implications for Secondary School Policymakers
School contexts are crucial in building up a professional learning community for the beginning Chinese-heritage language teachers to continuously develop assessment capabilities. The findings of this study show that beginning Chinese-heritage language teachers encountered difficulties connecting the knowledge and skills that they had learnt in teacher education within specific school contexts. They had expected there would be more opportunities to reflect on their own practices, and learn collaboratively with peers and from people who are more assessment literate, upon starting in the teaching profession. M. F. Hill (2011) found that PD workshops could effectively integrate assessment theories within specific school contexts, supporting teachers to bridge the gap between theories and own practices. Furthermore, principals, heads of department and mentor teachers are critical in

sustaining formative assessment practices in a specific school culture. Sadly, this did not appear to happen for these three beginning teachers. Therefore, it is important for school leaders and mentor teachers to consider the professional learning needs of new language teachers, especially around assessment. These beginning teachers needed more opportunities to extend their understandings about assessment. Cross-school communications were suggested by the participants in this study. Communications with teachers who have a similar cultural background may be helpful for the beginning teachers to reflect on the differences between New Zealand culture and Chinese culture, increasing awareness and understandings of formative assessment in the New Zealand context. To support Chinese beginning secondary language teachers' professional development, secondary school principals need to regard additional languages as an important part of the curriculum. East and Ward (2016) suggest secondary school principals make additional languages learning compulsory for Year 9 and 10 students, sending a positive signal about the value of additional language learning in school communities.

Implications for Beginning Chinese-Heritage Language Teachers

One of the implications for Chinese-heritage language teachers is that they need to be aware of different purposes and processes of assessment, particularly with respect to the process of interactive formative assessment. Beginning Chinese-heritage language teachers could reflect about the differences between their prior and current experience of assessment in different assessment contexts during the ITE programme and as they begin teaching. In this way, they could gain insights about advocated assessment practice in the New Zealand context and what they are good at in assessment and what they need to improve in future.

Furthermore, Farrell (2006) notes that teacher education may not prepare beginning teachers successfully, as "learning to teach is a highly situated, highly interpretive, and idiosyncratic activity" (p. 211). In other words, contextual factors and personal factors play important roles in beginning teachers' development of assessment as they start teaching. Thus, the other implication for beginning Chinese language teachers is to take the initiative themselves to learn about and to use assessment through attending professional development programmes and conferences regarding assessment within and across schools. In this way, beginning Chinese-heritage language teachers could develop assessment capabilities continuously.

Limitations of this Study

As with all research, limitations exist regarding the nature of this longitudinal case study. First, the approach of a case study, with limited numbers of participants, constrains the ability to generalise from this study. It may be problematic to apply the findings from this study to other educational contexts or other ethnic groups in the New Zealand context. However, as Miles et al. (2014) indicate, although case studies are designed to show a slice of a complex world, the interpretative approach can, and did, provide, rich and in-depth data about participants' practices.

In this sequential two-phase study, participant numbers in Phase 2 study were constrained by the numbers in the Phase 1 study. All the preservice Chinese-heritage language teachers in the Phase 1 study were invited to continue participating in the Phase 2 investigation. However, it was challenging for the cohort of participants to find suitable jobs after graduation in New Zealand secondary schools, and only three participants were successfully recruited for the Phase 2 study. Furthermore, these three participants were teaching in either private or integrated schools. The reality of this situation limited the intention to observe the changes of beginning teachers' conceptions and practices in a broader range of public, private and integrated schools, and thus, how beginning Chinese language teachers conceive and practise assessment in public schools remains unknown in this study. Also, although the three beginning Chinese-heritage language teachers started to teach in New Zealand secondary schools, all of them were assigned to teach the Chinese language in Year 9, as well as other subjects such as Chinese literature or ESOL for Years 12 and 13. Thus, the Year 9 Chinese language classes of the three beginning teachers were chosen to reduce the possible influences from different disciplines and year levels, even though the impact from high-stakes examinations in secondary schools may be less evident in classes below Year 11.

The investigation in this study focused on the beginning Chinese-heritage language teachers' experience of assessment, and thus the relatability and validity of data may be compromised as assessment conceptions from teacher educators, students and schools were not investigated. However, this longitudinal study intended to explore beginning Chinese-heritage language teachers' personal experience of assessment; detailed descriptions and key quotations from interviews and observations could assist readers to make decisions about what findings can be referenced in their contexts (Edwards, 2017).

A further limitation may be that the Phase 1 study was conducted in the English language, as the cohort of preservice teachers was at the exit of their ITE programme and thus seemed confident to use English during the interview. Nevertheless, the Phase 2 study was conducted in Chinese language based on suggestions from the three participants. Thus, the translation work was mainly conducted by me. As a second-language learner, my translation skills may not truly reflect what the participants were trying to express, and thus discussions about appropriate translations from the Chinese language to the English language frequently happened between my supervisors and myself, and participants and myself to increase the reliability of data.

Suggestions for Future Studies

The stories of Chinese-heritage language teachers' assessment learning from the preservice stage to in-service stage suggest the powerful influences of CHCs on their development of assessment capabilities in the New Zealand context. As teachers from China increasingly migrate to be teachers of Chinese language in New Zealand, more studies are recommended to explore teachers' continuous attempts to reconcile these two different cultures of pedagogy and assessment into the classroom. Furthermore, more studies are also recommended to investigate effective methods, such as teaching as inquiry, that ITE programmes could adopt to help preservice Chinese-heritage language teachers learn assessment through reflecting on the differences and connections between CHCs and local beliefs and practices.

The investigation of this study focused on beginning teachers' experiences. Future studies are suggested to include the perspectives of other stakeholders such as students, teacher educators, and headteachers/principals in the discussions of beginning teachers' professional development in assessment to understand such topics comprehensively.

As regards data-collection methods, further research is suggested to design observation schedules based on learning units rather than dispersed sessions. Classroom observations in this study were conducted once at each stage of investigation; each observation continued within one or two teaching sessions. However, assessment happens as a continuum; teachers' assessment decisions and further actions may not be observed in the observed sessions. Therefore, extending the length of time in classrooms would be helpful to gain more information regarding teachers' use of assessment and assessment information. Furthermore, the conceptual framework of this study includes forms, processes and purposes of assessment activities, and it was useful to describe and understand the beginning teachers' complicated

practices of assessment. Future studies are recommended to test this conceptual framework in other contexts, such as in other subjects and in other countries.

Conclusion

DeLuca, Willis, et al. (2019) suggest beginning teachers are found to be particularly underprepared for assessment in schools and make calls for more assessment-capable teachers. This longitudinal study traced changes in beginning Chinese-heritage language teachers' assessment conceptions and practice from preservice stage to in-service stage in the New Zealand context. At the exit of the 1-year ITE programme, the cohort of preservice Chinese-heritage language teachers had deepened their understandings of assessment and gained some confidence to use assessment in practice from the 1-year ITE programme. However, as they began teaching in New Zealand secondary schools, designing and implementing assessment activities in classrooms was found to be very challenging for the beginning Chinese-heritage teachers when they had limited understanding of the real purposes of formative assessment and, in particular, did not understand or recognise unplanned/interactive formative assessment. Furthermore, the findings show that the beginning Chinese-heritage language teachers may have made less use of summative assessment results to support student learning than they could have as they lacked skills and awareness of how to integrate summative and formative assessment into classroom practice. Five main conclusions of this study and its contributions are presented below.

First, this study demonstrates that beginning Chinese-heritage language teachers in New Zealand need careful preparation for assessment in ITE and continuous support to reflect on their own assessment practice and develop assessment capabilities when they begin teaching. In the present study, the 1-year ITE programme was found effective in changing the preservice Chinese-heritage language teachers' prior conceptions of assessment and equipping them with a beginning level of confidence to use assessment in practice. However, the limited assessment learning opportunities may not have been sufficient to support the beginning Chinese-heritage language teachers to reflect and refine their assessment practice as they began teaching in New Zealand secondary schools. This suggests the importance of providing a dedicated assessment course and direct assessment teaching during ITE programmes in order to support preservice Chinese-heritage language teachers to learn about and to use assessment and reflect upon their prior experience of assessment in the Chinese context.

Second, the findings of this study show that beginning Chinese-heritage language teachers' assessment learning is closely related to specific school contexts. In Phase 2, secondary schools' priorities of student achievement in qualifications, students' subject choice beyond Year 10 and school accountability impacted the beginning Chinese language teachers who put emphasis on using students' assessment feedback to improve their own teaching practice rather than to improve student learning. Furthermore, the beginning Chinese-heritage language teachers' assessment practice is also constrained by limited time to create assessment activities and their capability to use assessment in practice. Thus, support from their school, and colleagues within and beyond their school, is important in order to support formative assessment learning within New Zealand secondary schools.

Third, the findings of the current study could be of value in understanding the influence of CHCs on beginning Chinese-heritage language teachers' assessment conceptions and practice in New Zealand secondary schools. The beginning Chinese-heritage language teachers' prior experience of learning and/or teaching in CHCs, in which teachers are highly responsible for student-learning outcomes and responsive to expectations of school communities, impacted their teaching and assessment practice in New Zealand secondary schools. Furthermore, influenced by prior experience in CHCs in which a behaviouristic approach is adopted to improve teaching and learning, the beginning Chinese-heritage language teachers were challenged when creating teaching and assessment resources. Based on this understanding, more support within and across schools is suggested to encourage beginning Chinese-heritage language teachers to reflect upon the differences between their prior and current assessment experience and take initiatives in teaching and assessment practice in the New Zealand context.

Fourth, the described conceptual framework in the current study was useful to identify and understand teachers' complex assessment activities from three different perspectives (processes, purposes and forms of assessment), particularly when assessment is embedded in teaching or learning activities, and thus it has potential for use in other studies.

Finally, this longitudinal study confirms and extends previous literature regarding beginning Chinese-heritage language teachers' assessment learning experiences. The 2-year longitudinal dimension provided rich details regarding the beginning Chinese-heritage language teachers' conceptions and practices within and across preservice and beginning teaching, and thus was helpful to understand their professional development of assessment in their early teaching careers.

References

Aarts, R., Kools, Q., & Schildwacht, R. (2020). Providing a good start. Concerns of beginning secondary school teachers and support provided. *European Journal of Teacher Education*, *43*(2), 277-295. https://doi.org/10.1080/02619768.2019.1693992

Absolum, M., Flockton, L., Hattie, J., Hipkins, R., & Reid, I. (2009). *Directions for assessment in New Zealand: Developing students' assessment capabilities.* Paper presented at the International Symposium on Assessment for Learning. Queenstown, New Zealand.

Al Sawafi, O. S. (2014). *Investigating English teachers' beliefs and practices in relation to the Continuous Assessment reform in the Sultanate of Oman* (Doctoral dissertation, University of Leeds). http://etheses.whiterose.ac.uk/8101/1/Omar%20PhD%20thesis%20FINAL.pdf

Alonzo, D. (2016). *Development and application of a teacher assessment for learning (AfL) literacy tool*. [Unpublished doctoral dissertation]. University of New South Wales.

Angen, M. J. (2000). Evaluating interpretive inquiry: Reviewing the validity debate and opening the dialogue. *Qualitative Health Research, 10*(3), 378–395. https://doi.org/10.1177/104973230001000308

Araceli Ruiz-Primo, M., & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment, 11*(3–4), 237–263. https://doi.org/10.1080/10627197.2006.9652991

Ashton, K. (2018). Exploring teacher views of multi-level language classes in New Zealand secondary schools. *Teaching and Teacher Education, 69*, 104–118. https://doi.org/10.1016/j.tate.2017.10.002

Assessment Reform Group. (1999). *Assessment for learning: Beyond the black box*. Qualifications and Curriculum Authority.

Assessment Reform Group. (2002). *Assessment for learning: 10 principles*. http://www.aaia.org.uk/content/uploads/2010/06/Assessment-for-Learning-10-principles.pdf

Baker, C., & Jones, S. P. (1998). *Encyclopedia of bilingualism and bilingual education*. Multilingual Matters.

Barnes, N., Fives, H., & Dacey, C. M. (2015). Teachers' beliefs about assessment. In H. Fives & M. Gregoire Gill (Eds.), *International Handbook of Research on Teachers' Beliefs* (pp. 284–300). Routledge.

Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The Qualitative Report, 13*(4), 544–559.

Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education, 85*(5), 536–553. https://doi.org/10.1002/sce.1022

Bell, B., & Gilbert, J. K. (1996). *Teacher development: A model from science education.* Psychology Press.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5–25. https://doi.org/10.1080/0969594X.2010.513678

Bianco, J. L. (2011). Chinese: The gigantic up-and-comer. *Teaching and Learning Chinese in Global Contexts.* Continuum.

Black, P., Harrison, C., Lee, C., Marshall, B., & William, D. (2003). Formative and summative assessment: Can they serve learning together? *AERA Chicago.* http://www.kcl.ac.uk//depsta/education/hpages/pblackpubs.html.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (2006). The reliability of assessments. In J. Gardner (Ed.), *Assessment and learning* (pp. 119–131). Sage.

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice, 25*(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807

Boodhoo, C. (2018). *Assessment for learning in design and technology: An ethnographic study in Mauritius state secondary schools* [Doctoral dissertation, The University of Waikato]. https://hdl.handle.net/10289/12092

Booth, B. A. (2020). *Building assessment capability* [Unpublished doctoral dissertation], University of Auckland.

Booth, B., Hill, M. F., & Dixon, H. (2014). The assessment-capable teacher: Are we all on the same page. *Assessment Matters, 6*, 137–157.

Borg, S. (2006). *Teacher cognition and language education: Research and practice*. Continuum.

Borg, S. (2015). *Teacher cognition and language education: Research and practice* (2nd ed.). Bloomsbury Publishing.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative research journal*, *9*(2), 27-40. http://dx.doi:10.3316/QRJ0902027

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn*. National Academy Press.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.

Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. Sage.

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, *11*(4), 589-597. https://doi.org/10.1080/2159676X.2019.1628806

Braun, V., Clarke, V., Hayfield, N., & Terry, G. (2018). Thematic analysis. In P. Liamputtong (Ed.), *Handbook of research methods in health social sciences* (pp.843–860). Springer.

Brighouse, T., & Woods D. (1999). *How to improve your school*. Routledge.

Broadfoot, P. (2007). *An introduction to assessment*. Continuum Intl Pub Group.

Brookhart, S. M. (2001). *The "standards" and classroom assessment research*. Paper presented at the annual meeting of the American Association of Colleges for Teacher Education, Dallas. https://eric.ed.gov/?id=ED451189

Brookhart, S. M. (2002). What will teachers know about assessment, and how will that improve instruction. In R. W. Lizzitz & W. D. Schafer (Eds.), *Assessment in educational reform: Both means and ends* (pp. 2–17). Allyn & Bacon.

Brown, G. T. (2002). *Teachers' conceptions of assessment* [Doctoral dissertation, University of Auckland]. https://researchspace.auckland.ac.nz/handle/2292/63

Brown, G. T. (2003). *Teachers' instructional conceptions: Assessment's relationship to learning, teaching, curriculum, and teacher efficacy*. Paper presented at the Joint Conference of the Australian and New Zealand Associations for Research in Education (AARE/NZARE), Auckland, 28

Brown, G. T. (2011). New Zealand prospective teacher conceptions of assessment and academic performance: Neither student nor practicing teacher. In R. Kahn, J. C. McDermott, & A, Akimjak (Eds.), *Democratic access to education*. (pp. 119–132). Antioch University

Brown, G. T., & Gao, L. (2015). Chinese teachers' conceptions of assessment for and of learning: Six competing and complementary purposes. *Cogent Education, 2*(1), 993836. http://dx.doi.org/10.1080/2331186X.2014.993836

Brown, G. T., Kennedy, K. J., Fok, P. K., Chan, J. K., & Yu, W. M. (2009). Assessment for student improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy & Practice, 16*(3), 347–363. https://doi.org/10.1080/09695940903319737

Brown, G. T., & Remesal, A. (2012). Prospective teachers' conceptions of assessment: A cross-cultural comparison. *The Spanish Journal of Psychology, 15*(01), 75–89. https://doi.org/10.5209/rev_SJOP.2012.v15.n1.37286

Bullough, R. V. (1997). Becoming a teacher: Self and the social location of teacher education. In B. Biddle, T. Good, & I. Goodson (Eds.), *International handbook of teachers and teaching* (pp. 79–134). Kluwer Academic Publishers. https://doi.org/10.1007/978-94-011-4942-6_3

Büyükkarci, K. (2014). Assessment Beliefs and Practices of Language Teachers in Primary Education. *International Journal of Instruction*, *7*(1), 107-120. https://eric.ed.gov/?id=EJ1085246

Cameron, M. (2007). *Learning to teach: A literature review on induction theory and practice*. New Zealand Teachers Council, Wellington.

Candy, P. C. (1989). Alternative paradigms in educational research. *The Australian Educational Researcher, 16*(3), 1–11.

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. (2012). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings*. Routledge.

Carless, D., & Lam, R. (2014). Developing assessment for productive learning in Confucian-influenced settings. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning* (pp. 167–179). Springer.

Chandan, S. R. (2018). *Assessment of yield potential and water productivity of hybrid maize under drip fertigation* [Doctoral dissertation], Vasantrao Naik Marathwada Krishi Vidyapeeth, Parbhani. https://krishikosh.egranth.ac.in/handle/1/5810069676

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Sage.

Chen, J. (2015). Teachers' conceptions of approaches to teaching: A Chinese perspective. *The Asia-Pacific Education Researcher*, *24*(2), 341–351. https://doi.org/10.1007/s40299-014-0184-3

Chen, J., & Brown, G. T. (2016). Tensions between knowledge transmission and student-focused teaching approaches to assessment purposes: Helping students improve through transmission. *Teachers and Teaching, 22*(3), 350–367. https://doi.org/10.1080/13540602.2015.1058592

Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures. *Language Testing, 21*(3), 360–389. https://doi.org/10.1191/0265532204lt288oa

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review, 24*(2), 205–249. https://doi.org/10.1007/s10648-011-9191-6

Cohen, L., Manion, L., Morrison, K. (2007). Research methods in education (6th ed.). Routledge.

Coombs, A., DeLuca, C., LaPointe-McEwan, D., & Chalas, A. (2018). Changing approaches to classroom assessment: An empirical study across teacher career stages. *Teaching and Teacher Education, 71*, 134–144. https://doi.org/10.1016/j.tate.2017.12.010

Coombs, A., DeLuca, C., & MacGregor, S. (2020). A person-centered analysis of teacher candidates' approaches to assessment. *Teaching and Teacher Education, 87*, 102952. https://doi.org/10.1016/j.tate.2019.102952

Cooper, B., & Cowie, B. (2010). Collaborative research for assessment for learning. *Teaching and Teacher Education, 26*(4), 979–986. https://doi.org/10.1016/j.tate.2009.10.040

Corbin, J., & Strauss, A. (2008). Strategies for qualitative data analysis. *Basics of Qualitative Research. Techniques and Procedures for Developing Grounded Theory*, *3*, 65–86. https://dx.doi.org/10.4135/9781452230153.n4

Corcoran, C. (2014). Chinese learning styles: Blending Confucian and western theories. *Journal of Instructional Pedagogies, 13*.1–10. https://files.eric.ed.gov/fulltext/EJ1060088.pdf

Cortazzi, M., & Jin, L. (1996). Cultures of learning: Language classrooms in China. In H. Coleman (Ed.), *Society and the language classroom* (pp. 169-206). https://doi.org/10.1017/S0261444800008351

Cowie, B., & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education: Principles, Policy & Practice, 6*(1), 101–116. https://doi.org/10.1080/09695949993026

Creswell, J. W. (2007). *Qualitative enquiry and research design: Choosing among five approaches*. Sage.

Creswell, J. W. (2011). Controversies in mixed methods research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (4th ed., pp. 269–284). Sage.

Creswell, J. W. & Poth, C. N. (2018). *Qualitative inquiry & research design: Choosing among five approaches* (4th ed.). Sage.

Crooks, T. (2001). *The validity of formative assessments.* [Paper presentation]. British Educational Research Association Annual Conference, University of Leeds, 13–15.

Crooks, T. (2011). Assessment for learning in the accountability era: New Zealand. *Studies in Educational Evaluation, 37*(1), 71–77. https://doi.org/10.1016/j.stueduc.2011.03.002

Darling-Hammond, L., Hammerness, K., Grossman, P., Rust, F., & Shulman, L. (2005). The design of teacher education programs. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 390–441). Jossey Bass.

Darling-Hammond, L., & Sclan, E. M. (1996). Who teaches and why: Dilemmas of building a profession for twenty-first century schools. In J. P. Sikula, T. J. Buttery, & E. Guyton (Eds.), *Handbook of research on teacher education* (2nd ed., pp. 67–101). Routledge.

Daugherty, R., & Ecclestone, K. (2006). Constructing assessment for learning in the UK policy environment In J. Gardner (Ed.), *Assessment and learning* (pp. 149–167). Sage.

Davey, G., De Lian, C., & Higgins, L. (2007). The university entrance examination system in China. *Journal of further and Higher Education*, *31*(4), 385–396. https://doi.org/10.1080/03098770701625761

Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing, 21*(3), 305–334. https://doi.org/10.1191/0265532204lt286oa

Davison, C. (2019). Using assessment to enhance learning in English language education. In X. Gao (Ed.), *Second handbook of English language teaching* (pp. 433–454). https://doi.org/10.1007/978-3-030-02899-2_21

Davison, C., & Gao, X. (2019). *Using assessment to enhance learning in English language education.* https://doi.org/10.1007/978-3-030-02899-2_21

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly, 43*(3), 393–415. https://doi.org/10.1002/j.1545-7249.2009.tb00242.x

Dello-Iacovo, B. (2009). Curriculum reform and 'quality education' in China: An overview. *International Journal of Educational Development, 29*(3), 241–249. https://doi.org/10.1016/j.ijedudev.2008.02.008

DeLuca, C., Bellara, A. (2013). The current state of assessment education: Aligning policy, standards, and teacher education curriculum. *Journal of Teacher Education*, *64*(4), 356–372. https://doi.org/10.1177/0022487113488144

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., Coombs, A., & LaPointe-McEwan, D. (2019). Assessment mindset: Exploring the relationship between teacher mindset and approaches to classroom assessment. *Studies in Educational Evaluation, 61*, 159–169. https://doi.org/10.1016/j.stueduc.2019.03.012

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/10.1080/0969594X.2010.516643

DeLuca, C., Luu, K., Sun, Y., & Klinger, D. A. (2012). Assessment for learning in the classroom: Barriers to implementation and possibilities for teacher professional learning. *Assessment Matters, 4*, 5–29.

DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D., & Luhanga, U. (2018). Teachers' approaches to classroom assessment: A large-scale survey. *Assessment in Education: Principles, Policy & Practice, 25*(4), 355–375. https://doi.org/10.1080/0969594X.2016.1244514

DeLuca, C., Willis, J., Cowie, B., Harrison, C., Coombs, A., Gibson, A., & Trask, S. (2019). Policies, programs, and practices: Exploring the complex dynamics of assessment education in teacher education across four countries. *Frontiers in Education, 4*. https://doi.org/10.3389/feduc.2019.00132

Denzin, N. K., & Lincoln, Y. S. (2017). Introduction: The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (5th ed.). Sage.

Dixon, H., & Hawe, E. (2018). Developing assessment-capable teachers through engagement. *Teacher Learning with Classroom Assessment: Perspectives from Asia Pacific, 59*. https://doi.org/10.1007/978-981-10-9053-0_4

Earl, L. (2003). *Assessment as learning: Using classroom assessment to maximise student learning.* Corwin Press.

East, M. (2008). Language evaluation policies and the use of support resources in assessments of language proficiency. *Current Issues in Language Planning, 9*(3), 249–261. https://doi.org/10.1080/14664200802139539

East, M. (2012). Working towards a motivational pedagogy for school programmes in additional languages. *Curriculum Matters, 8*, 128–148.

East, M. (2015). Coming to terms with innovative high-stakes assessment practice: Teachers' viewpoints on assessment reform. *Language Testing, 32*(1), 101–120. https://doi.org/10.1177/0265532214544393

East, M. (2016a). *Assessing foreign language students' spoken proficiency: Stakeholder perspectives on assessment innovation.* Springer. https://doi.org/10.1007/978-981-10-0303-5

East, M. (2016b). *Nomination for Associate Professor Martin East Tertiary Teaching Excellence Awards.* Faculty of Education and Social Work. The University of Auckland. https://cdn.auckland.ac.nz/assets/auckland/creative/about-the-faculty/documents/martin-east-portfolio.pdf

Edwards, F. (2017). *The development of summative assessment literacy: An exploration of the experiences of beginner secondary science teachers in New Zealand* [Doctoral dissertation, The University of Waikato]. https://researchcommons.waikato.ac.nz/handle/10289/11567

Edwards, F., & Edwards, R. (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

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

Ellis, R. (2012). *Language teaching research and language pedagogy.* John Wiley & Sons.

Ellis, R. (2012). *Language teaching research and language pedagogy*. John Wiley & Sons.

Engelsen, K. S., & Smith, K. (2014). Assessment literacy. In K. Engelsen & K. Smith (Eds.), *Designing assessment for quality learning* (pp. 91–107). Springer. https://doi.org/10.1007/978-94-007-5902-2_6

Eyers, G. (2014). *Preservice teachers' assessment learning: Change, development and growth* (Doctoral dissertation, University of Auckland). https://researchspace.auckland.ac.nz/handle/2292/23757.

Eyers, G., & Hill, M. (2004). Improving student learning? Research evidence about teacher feedback for improvement in New Zealand schools. *Waikato Journal of Education, 10.* https://doi.org/10.15663/wje.v10i1.345

Farrell, T. S. (2006). The first year of language teaching: Imposing order. *System*, *34*(2), 211–221. https://doi.org/10.1016/j.system.2005.12.001

Farrell, T. S. (2012). Novice-service language teacher development: Bridging the gap between preservice and in-service education and development. *Tesol Quarterly, 46*(3), 435–449. https://doi.org/10.1002/tesq.36

Fenstermacher, G. D., & Berliner, D. C. (1983). *A conceptual framework for the analysis of staff development*. Rand.

Flick, U. (2014). *An introduction to qualitative research* (5th ed.). Sage.

Flockton, L. (2012). Commentary: Directions for assessment in New Zealand. *Assessment Matters, 4*, 129.

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/10.1016/j.tate.2005.09.002

Fong, R. (2007). Intercultural communication: Chinese culture in UK education. In M. N. Craith (Ed), *Language, power and identity politics* (pp. 146–166). Palgrave Macmillan. https://doi.org/10.1057/9780230592841_8

Fullan, M. (2005). Turnaround leadership. *The Educational Forum*, *69*(2), 174–181. https://doi.org/10.1080/00131720508984681

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, *9*(2), 113-132. https://doi.org/10.1080/15434303.2011.642041

Garbett, D. (2011). Constructivism deconstructed in science teacher education. *Australian Journal of Teacher Education, 36*(6), 3. http://dx.doi.org/10.14221/ajte.2011v36n6.5

Gardner, J. (2006). Assessment for learning: A compelling conceptualization In J. Gardner (Ed.), *Assessment and learning* (pp. 197–204). Sage.

Gardner, J., Harlen, W., & Hayward, L. (2010). *Developing teacher assessment*. McGraw-Hill Education.

Garrison, C., & Ehringhaus, M. (2007). Formative and summative assessments in classroom. In C. Garrisson, D. Chandler, & M. Eringhaus (Eds.), *Effective classroom assessment* (pp. 1–3). Westerville.

Gipps, C., Brown, M., McCallum, B., & McAlister, S. (1995). *Intuition or evidence. Teachers and national assessment of seven-year-olds.* Open University Press.

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

Grudnoff, L. (2012). All's well? New Zealand beginning teachers' experience of induction provision in their first six months in school. *Professional Development in Education, 38*(3), 471–485. https://doi.org/10.1080/19415257.2011.636894

Gu, P. Y. (2014). The unbearable lightness of the curriculum: What drives the assessment practices of a teacher of English as a foreign language in a Chinese secondary school? *Assessment in Education: Principles, Policy & Practice, 21*(3), 286–305. https://doi.org/10.1080/0969594X.2013.836076

Guan, Q., & Meng, W. (2007). China's new national curriculum reform: Innovation, challenges and strategies. *Frontiers of Education in China, 2*(4), 579–604. https://doi.org/10.1007/s11516-007-0043-6

Gustafsson, J. (2017). *Single case studies vs. multiple case studies: A comparative study.* [Doctoral thesis, Halmstad University].

http://hh.diva-portal.org/smash/record.jsf?pid=diva2%3A1064378&dswid=-1102

Hammerness, K., Darling-Hammond, L., & Bransford, J. (2005). How teachers learn and develop. L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a*

changing world: What teachers should learn and be able to do (pp. 358–389). Jossey-Bass.

Hara, K. (2011). *Educational change beyond borders: International Baccalaureate in New Zealand* [Doctoral dissertation, Auckland University of Technology]. https://openrepository.aut.ac.nz/handle/10292/3262

Hargreaves, A., Earl, L., & Schmidt, M. (2002). Perspectives on alternative assessment reform. *American Educational Research Journal, 39*(1), 69–95. https://doi.org/10.3102/00028312039001069

Hargreaves, E. (2005). Assessment for learning? thinking outside the (black) box. *Cambridge Journal of Education, 35*(2), 213–224. https://doi.org/10.1080/03057640500146880

Hargreaves, E. (2007). The validity of collaborative assessment for learning. *Assessment in Education*, *14*(2), 185-199. https://doi.org/10.1080/09695940701478594

Harlen, W. (2007). *Assessment of learning*. Sage.

Harlen, W. (2013). *Assessment & inquiry-based science education: Issues in policy and practice*. Global Network of Science Academies.

Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice, 4*(3), 365–379. https://doi.org/10.1080/0969594970040304

Harrell, M. C., & Bradley, M. A. (2009). *Data collection methods. Semi-structured interviews and focus groups*. Rand National Defense Research Inst.

Hatch, J. A. (2002). *Doing qualitative research in education settings*. Suny Press.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Haverinen-Shaughnessy, U., Moschandreas, D. J., & Shaughnessy, R. J. (2011). Association between substandard classroom ventilation rates and students' academic achievement. *Indoor Air*, *21*(2), 121–131. https://doi.org/10.1080/03098770701625761

Hawe, E. M. (2000). *Assessment in a pre-service teacher education programme: The rhetoric and the practice* [Doctoral dissertation, University of Auckland]. https://researchspace.auckland.ac.nz/handle/2292/632

Heflich, D. A., & Rice, M. L. (1999). Online survey research: A venue for reflective conversation and professional development. In J. Price, J. Willis, D. Willis, M. Jost & S. Boger-Mehall (Eds.), *Society for Information Technology & Teacher Education International Conference* (pp. 1350–1355). https://www.learntechlib.org/primary/p/7754/

Hill, K., & McNamara, T. (2012). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing, 29*(3), 395–420. https://doi.org/10.1177/0265532211428317

Hill, M. F. (2000). *Remapping the assessment landscape: Primary teachers reconstructing assessment in self-managing schools* [Unpublished doctoral dissertation]. University of Waikato.

Hill, M. F. (2011). 'Getting traction': Enablers and barriers to implementing assessment for learning in secondary schools. *Assessment in Education: Principles, Policy & Practice, 18*(4), 347–364. https://doi.org/10.1080/0969594X.2011.600247

Hill, M. F., Cowie, B., Gilmore, A., & Smith, L. F. (2010). Preparing assessment-capable teachers: What should preservice teachers know and be able to do? *Assessment Matters, 2*, 43.

Hill, M. F., Ell, F. R., & Eyers, G. (2017, May). Assessment capability and student self-regulation: The challenge of preparing teachers. *Frontiers in Education 2*, 1–21. https://doi.org/10.3389/feduc.2017.00021

Hill, M. F., Ell, F., Grudnoff, L., & Limbrick, L. (2014). Practise what you preach: Initial teacher education students learning about assessment. *Assessment Matters*, *7*(90), e112.

Hill, M. F., & Eyers, G. E. (2016). Moving from student to teacher: changing perspectives about assessment through teacher education. In S. M. Brookhart (Ed.), *Handbook of human and social conditions in assessment* (pp 57–76). https://doi.org/10.3389/feduc.2017.00021

Hill, M. F., Gunn, A., Cowie, B., Smith, L. F., & Gilmore, A. (2014). Preparing primary and early childhood initial teacher education students to use assessment in teaching. *Assessment Matters*, *7*, 4–23.

Hill, M. F., Tan, K., & Jiang, H. (2018). Classroom Assessment for Teacher Learning and Student Learning. In H. Jiang & M. F. Hill (Eds.), *Teacher learning with classroom assessment* (pp. 193–201). Springer. https://doi.org/10.1007/978-981-10-9053-0_12

Ho, D. G. (2006). The focus group interview: Rising to the challenge in qualitative research methodology. *Australian Review of Applied Linguistics, 29*(1), 5.1–5.19. https://doi.org/10.2104/aral0605

Hoepfl, M. C. (1997). Choosing qualitative research: A primer for technology education researchers. *Journal of Technology Education, 9*(1), 47–63.

Inbar-Lourie, O., & Donitsa-Schmidt, S. (2009). Exploring classroom assessment practices: The case of teachers of English as a foreign language. *Assessment in Education: Principles, Policy & Practice*, *16*(2), 185–204. https://doi.org/10.1080/09695940903075958

International Baccalaureate Organization. (2010). *IB diploma examiners*. Switzerland. www.ibo.org/examiners

James, M. (2006). Assessment, teaching and theories of learning. *Assessment and Learning, 47*, 60.

Jiang, H. (2015). *Learning to teach with assessment: A student teaching experience in China*. Springer. https://doi.org/10.1007/978-981-287-272-2

Jiang, H., & Hill, M. F. (2018). *Teacher learning with classroom assessment: Perspectives from Asia Pacific*. Springer. https://doi.org/10.1007/978-981-10-9053-0

Jones, C. (2014). The shortage of students studying languages for NCEA Level 3. *Set*, *2*, 24–32. https://www.nzcer.org.nz/system/files/set2014_2_024.pdf

Jones, J., & Webb, M. (2006). *Assessment for learning (AfL) across the school: A case study in whole school capacity building.* Paper presented at the British Educational Research Association Annual Conference, University of Warwick, 6–9 September 2006. http://www.leeds.ac.uk/educol/documents/160634.htm

Jordan, B., & Putz, P. (2004). Assessment as practice: Notes on measures, tests, and targets. *Human Organization, 63*(3), 346–358. https://doi.org/10.17730/humo.63.3.yj2w5y9tmblc422k

Kai, T. (2015). *The effectiveness and value of the Mandarin Language Assistant programme in Auckland–a case study* [Doctoral dissertation, University of Auckland]. https://researchspace.auckland.ac.nz/handle/2292/26533

Karp, G. G., & Woods, M. L. (2008). Preservice teachers' perceptions about assessment and its implementation. *Journal of Teaching in Physical Education, 27*(3), 327–346. https://doi.org/10.1123/jtpe.27.3.327

Kennedy, K. J. (2016). Exploring the influence of culture on assessment: The case of teachers' conceptions of assessment in Confucian Heritage Cultures. In G. T. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 404–419). Routledge

Kitzinger, J. (1995). Qualitative research. Introducing focus groups. *BMJ (Clinical Research Ed.), 311*(7000), 299–302. https://doi.org/10.1136/bmj.311.7000.299

Koh, K., Burke, L. E. C. A., Luke, A., Gong, W., & Tan, C. (2018). Developing the assessment literacy of teachers in Chinese language classrooms: A focus on assessment task design. *Language Teaching Research*, *22*(3), 264–288. https://doi.org/10.1177%2F1362168816684366

Knight, P. T. (2002). Summative assessment in higher education: Practices in disarray. *Studies in Higher Education, 27*(3), 275–286. https://doi.org/10.1080/03075070220000662

Kvale, S., & Brinkmann, S. (2009). *Interviews: Learning the craft of qualitative research interviewing*. Sage.

Lacey, A., & Luff, D. (2001). Theories and methods in qualitative data analysis. *Qualitative data analysis* (pp. 320–357). Trent Focus Sheffield.

Lam, R. (2016). Implementing assessment for learning in a Confucian context: The case of Hong Kong 2004–14. In D. Wyse, L. Hayward & J. Pandya (Eds.), *The Sage handbook of curriculum, pedagogy and assessment* (2nd ed., pp. 756–771). Sage.

Langdon, F. (2011). Shifting perception and practice: New Zealand beginning teacher induction and mentoring as a pathway to expertise. *Professional Development in Education*, *37*(2), 241–258. https://doi.org/10.1080/19415257.2010.509658

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

Lederman, L. C. (1990). Assessing educational effectiveness: The focus group interview as a technique for data collection. *Communication Education*, *39*(2), 117–127. https://doi.org/10.1080/03634529009378794

Lee, C., & Wiliam, D. (2005). Studying changes in the practice of two teachers developing assessment for learning. *Teacher Development, 9*(2), 265–283. https://doi.org/10.1080/13664530500200244

Leung, C., & Lewkowicz, J. (2006). Expanding horizons and unresolved conundrums: Language testing and assessment. *Tesol Quarterly, 40*(1), 211–234. https://doi.org/10.2307/40264517

Leung, C., & Mohan, B. (2004). Teacher formative assessment and talk in classroom contexts: Assessment as discourse and assessment of discourse. *Language Testing, 21*(3), 335–359. https://doi.org/10.1191/0265532204lt287oa

Levy-Vered, A., & Nasser-Abu Alhija, F. (2015). Modelling beginning teachers' assessment literacy: The contribution of training, self-efficacy, and conceptions of assessment. *Educational Research and Evaluation, 21*(5–6), 378–406. https://doi.org/10.1080/13803611.2015.1117980

Lincoln, Y. S., & González y González, E. M. (2008). The search for emerging decolonizing methodologies in qualitative research: Further strategies for liberatory and democratic inquiry. *Qualitative Inquiry, 14*(5), 784–805. https://doi.org/10.1177/1077800408318304

Lincoln, Y. S., & Guba, E. (1985). Contextualization: Evidence from distributed teams. *Information Systems Research*, *16*(1), 9–27.

Litosseliti, L. (2003). *Using focus groups in research*. A&C Black.

Liu, J., Johnson, R., & Fan, X. (2016). A comparative study of Chinese and United States pre-service teachers' perceptions about ethical issues in classroom assessment. *Studies in Educational Evaluation, 48*, 57–66.

Lodico, M. G., Spaulding, D. T., & Voegtle, K. H. (2010). *Methods in educational research: From theory to practice*. John Wiley & Sons.

Lorente-Catalán, E., & Kirk, D. (2016). Student teachers' understanding and application of assessment for learning during a physical education teacher education course. *European Physical Education Review, 22*(1), 65–81. https://doi.org/10.1177/1356336X15590352

Mack, N., Woodsong, C., MacQueen, K. M., Guest, G., & Namey, E. (2005). *Qualitative research methods: A data collectors field guide*. Family Health International.

Mackenzie, N., & Knipe, S. (2006). Research dilemmas: Paradigms, methods and methodology. *Issues in Educational Research, 16*(2), 193–205.

Maclellan, E. (2001). Assessment for learning: The differing perceptions of tutors and students. *Assessment & Evaluation in Higher Education, 26*(4), 307–318. https://doi.org/10.1080/02602930120063466

Marshall, B., & Drummond, M. J. (2006). How teachers engage with assessment for learning: Lessons from the classroom. *Research Papers in Education, 21*(02), 133–149. https://doi.org/10.1080/02671520600615638

Maxwell, J. A. (2012). *Qualitative research design: An interactive approach*. Sage Publications.

May, S. (2002). Where to from here: Charting a way forward for language and education policy in Aotearoa/New Zealand. *The TESOLANZ Journal, 10*, 22–35.

McDonald, L. (2004). Effective mentoring of student teachers: Attitudes, characteristics and practices of successful associate teachers within a New Zealand context. *New Zealand Journal of Teachers' Work*, *1*(2), 85–94.

McMillan, J. H. (2008). *Assessment essentials for standards-based education*. Corwin Press.

McNamara, T., Hill, K., & May, L. (2002). 12. Discourse and assessment. *Annual Review of Applied Linguistics*, *22*, 221–242. https://doi.org/10.1017/S0267190502000120

McPhail, G. (2016). The fault lines of recontextualisation: The limits of constructivism in education. *British Educational Research Journal, 42*(2), 294–313. https://doi.org/10.1002/berj.3199

Merriam, S. B. (1998). *Qualitative research and case study applications in education.* Revised and Expanded from *Case study research in education*. Jossey-Bass Publishers.

Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. Wiley.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Sage.

Miller, T., Birch, M., Mauthner, M., & Jessop, J. (2012). *Ethics in qualitative research*. Sage.

Ministry of Education. (2007). *The New Zealand curriculum*. Learning Media

Ministry of Education. (2010). *Ministry of Education position paper: Assessment (schooling sector)*. Learning Media.

Ministry of Education China. (2010). Framework for national mid- to long-term educational reform and development planning (2010–2020). http://www.most.gov.cn/kjzc/gjkjzc/gjkjzczh/201308/P020130823574944373948.pdf

Morrell, P. D., & Carroll, J. B. (2010). *Conducting educational research: A primer for teachers and administrators*. Sense.

Moss, P. (2008). Sociocultural implications for assessment I: Classroom assessment. In P. A. Moss, D. Pullin, J. P. Gee, E. H. Haertel, & L. J. Young (Eds.)*, Assessment, equity, and opportunity to learn* (pp. 222–258). Cambridge University Press

Naruemon, D. (2013). *Thai pre-service teachers' beliefs about the learner-centred approach and their classroom practices* (Doctoral dissertation, Newcastle University). http://theses.ncl.ac.uk/jspui/handle/10443/2428

Nasr, M., Bagheri, M. S., & Sadighi, F. (2020). Iranian English language teachers' perceptions of monitoring and scaffolding practices of assessment for learning: A focus on gender and class size. *International Journal of Foreign Language Teaching and Research, 8*(29), 75–90.

New Zealand Qualification Authority. (2010). NZQA homepage. https://www.nzqa.govt.nz/

New Zealand Teachers Council. (2007). *Graduating teacher standards.* http://www.teacherscouncil.govt.nz/education/gts/index.stm

Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education, 14*(2), 149–170. https://doi.org/10.1080/09695940701478321

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

Northfield, J., & Gunstone, R. (1997). Teacher education as a process of developing teacher knowledge. In I. J. Loughran & T. Russel (Eds.), *Teaching about teaching: Purpose, passion and pedagogy in teacher education* (pp. 48–56). Falmer Press

Nusche, D., Laveault, D., MacBeath, J., & Santiago, P. (2012). *OECD reviews of evaluation and assessment in education: New Zealand.* https://www.udir.no/globalassets/upload/5/oecd_evaluering.pdf

Orton, J. (2011). Educating Chinese language teachers–Some fundamentals. In L. Tsung & K. Cruickshank (Eds.), *Teaching and learning Chinese in global contexts: CFL worldwide* (pp. 151–164). Bloomsbury

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

Palys, T. (2008). Purposive sampling. *The Sage Encyclopedia of Qualitative Research Methods, 2*(1), 697–698.

Patton, M. Q. (2014). *Qualitative research & evaluation methods: Integrating theory and practice*. Sage.

Perez, A., & Shin, M. H. (2016). Study on learning styles and Confucian culture. *Indian Journal of Science and Technology*, *9*(26), 1–5. https://doi.org/10.17485/ijst/2016/v9i26/97395

Phothongsunan, S. (2015). Causal attributions of English learning success of Thai students in an international university. *NIDA Development Journal, 55*(2), 101–129. https://doi.org/10.14456/ndj.2015.4

Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting L2 development*. Springer Science & Business Media.

Poole, A. (2016). 'Complex teaching realities' and 'deep rooted cultural traditions': Barriers to the implementation and internalisation of formative assessment in China. *Cogent Education, 3*(1), 1156242. http://dx.doi.org/10.1080/2331186X.2016.1156242

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice, 48*(1), 4–11. https://doi.org/10.1080/00405840802577536

Purpura, J. E., & Turner, C. E. (2018, July 2–6). A *learning-oriented assessment approach in classroom contexts: An expanded conceptualization of performance*. Workshop presented at LTRC, Auckland.

Quilter, S. M., & Gallini, J. K. (2000). Teachers' assessment literacy and attitudes. *The Teacher Educator, 36*(2), 115–131. https://doi.org/10.1080/08878730009555257

Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing, 18*(4), 429–462. https://doi.org/10.1177/026553220101800407

Rea-Dickins, P. (2006). Currents and eddies in the discourse of assessment: A learning-focused interpretation. *International Journal of Applied Linguistics, 16*(2), 163–188. https://doi.org/10.1111/j.1473-4192.2006.00112.x

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

Richardson, V. & Placier, P. (2001). Teacher change. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 905–947). Aera Books.

Rodriguez, J. (2019). *Teacher perceptions of the impact of a school-developed professional learning program on their effectiveness in assessing students' engagement and progress in learning* (Unpublished master's thesis). University of New South Wales.

Rubin, H. J., & Rubin, I. S. (2005). *Qualitative interviewing: The art of hearing data* (2nd ed.). Sage. https://dx.doi.org/10.4135/9781452226651

Ruiz-Primo, M. A. (2011). Informal formative assessment: The role of instructional dialogues in assessing students' learning. *Studies in Educational Evaluation, 37*(1), 15–24. https://doi.org/10.1016/j.stueduc.2011.04.003

Ruiz-Primo, M. A., & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment, 11*(3–4), 237–263.

Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching, 44*(1), 57–84. https://doi.org/10.1002/tea.20163

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119–144. https://doi.org/10.1007/BF00117714

Saito, H., & Inoi, S. (2017). Junior and senior high school EFL teachers' use of formative assessment: A mixed-methods study. *Language Assessment Quarterly, 14*(3), 213–233. https://doi.org/10.1080/15434303.2017.1351975

Saldaña, J. (2009). *The coding manual for qualitative researchers* (3rd ed.). Sage.

Schafer, W. D. (1993). Assessment literacy for teachers. *Theory into Practice, 32*(2), 118–126. https://doi.org/10.1080/00405849309543585

Schwandt, T. A., & Gates, E. F. (2017). Official science methodology is the study of how a particular kind of investigation should. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (5th ed., pp. 341–358). Sage.

Scriven, M. (1967). *The methodology of assessment.* American Educational Research Association.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4–14. https://doi.org/10.3102/0013189X029007004

Simons, H. (2009). *Case study research in practice*. SAGE Publications.

Smith, L. F., Hill, M. F., Cowie, B., & Gilmore, A. (2014). Preparing teachers to use the enabling power of assessment. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds), *Designing assessment for quality learning* (pp. 303–323). Springer. https://doi.org/10.1007/978-94-007-5902-2_19

Song, X. (2016). Fairness in educational assessment in China: Historical practices and contemporary challenges. In S. Scott, D. Scott, & C. Webber (Eds), *Assessment in education* (pp. 67–89). Springer. https://doi.org/10.1007/978-3-319-23398-7_4

Stake, R. E. (1995). *The art of case study research*. Sage.

Stiggins, R. J. (1999). Assessment, student confidence, and school success. *The Phi Delta Kappan, 81*(3), 191–198. https://www.jstor.org/stable/20439619

Stiggins, R. J. (2007). Educational leadership. *Assessment Through the Student's Eyes*, *64*(8), 22–26.

Sun, D. (2010). *Developing 'my way' in Chinese language teaching: Qualitative case studies of teachers' personal practical knowledge* (Doctoral thesis, Victoria University of Wellington). http://hdl.handle.net/10063/1480

Tam, A. C. (2013). Pre-service teachers' beliefs about the switch of medium-of-instruction and their influence on lesson planning and teaching practice. *Asia Pacific Journal of Education, 33*(4), 476–492. https://doi.org/10.1080/02188791.2013.809693

Tan, C. (2016). Tensions and challenges in China's education policy borrowing. *Educational Research, 58*(2), 195–206. https://doi.org/10.1080/00131881.2016.1165551

Tan, C. (2017). Confucianism and education. In G. Noblit (Ed.), *Oxford research encyclopedia of education*. Oxford University Press https://doi.org/10.1093/acrefore/9780190264093.013.226

Tan, C., & Chua, C. S. (2015). Education policy borrowing in China: Has the west wind overpowered the east wind? *Compare: A Journal of Comparative and International Education, 45*(5), 686–704. https://doi.org/10.1080/03057925.2013.871397

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. (2007). Assessment for learning: Understanding theory to improve practice. *Journal of Further and Higher Education, 31*(4), 363–371. https://doi.org/10.1080/03098770701625746

Teaching Council of Aotearoa New Zealand. (n.d.). *Getting Certificated.* https://teachingcouncil.nz/content/getting-started-%E2%80%93-certification

Terry, G., Hayfield, N., Clarke, V., & Braun, V. (2017). Thematic analysis. In C. Willig & W. Stainton-Rogers (Eds), *The Sage handbook of qualitative research in psychology* (pp. 17–37). Sage.

Thanh, N. C., & Thanh, T. (2015). The interconnection between interpretivist paradigm and qualitative methods in education. *American Journal of Educational Science, 1*(2), 24–27.

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation, 27*(2), 237–246. https://doi.org/10.1177/1098214005283748

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. 127–46). Macmillan

Thompson, M., & Wiliam, D. (2007). *Tight but loose: A conceptual framework for scaling up school reforms* [Paper presentation]. Annual Meeting of the American Educational Research Association. Chicago.

Tierney, R. D. (2006). Changing practices: Influences on classroom assessment. *Assessment in Education, 13*(3), 239–264. https://doi.org/10.1080/09695940601035387

Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher professional learning and development: Best evidence synthesis iteration*. Ministry of Education.

Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal, 27*(5), 615–631. https://doi.org/10.1080/01411920120095780

Tung, C. K. (2019). *A phenomenographic study of Singapore Chinese language teachers' experience of assessment tension* (Doctoral dissertation, Nanyang Technological University). http://5gsgedu.net/ppt/chenguangtong.pdf

Tunstall, P., & Gipps, C. (1996). 'How does your teacher help you to make your work better?' Children's understanding of formative assessment. *The Curriculum Journal, 7*(2), 185–203. https://doi.org/10.1080/0958517960070205

Turner, C., & Purpura, J. (2015). Learning-oriented assessment in second and foreign language classrooms. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 255–272). De Gruyter Mouton

Turner, D. W., III. (2010). Qualitative interview design: A practical guide for novice investigators. *The Qualitative Report, 15*(3), 754.

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.

Walsh, I., Holton, J. A., Bailyn, L., Fernandez, W., Levina, N., & Glaser, B. (2015). What grounded theory is a critically reflective conversation among scholars. *Organizational Research Methods, 18*(4), 581–599. https://doi.org/10.1177/1094428114565028

Ward, D., & East, M. (2016). Is language learning languishing? Stakeholders' perspectives on the provision of secondary school foreign language programmes in New Zealand. *New Zealand Language Teacher,* 42–44.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*(1), 3–14. https://doi.org/10.1016/j.stueduc.2011.03.001

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy & Practice, 11*(1), 49–65. https://doi.org/10.1080/0969594042000208994

Wilkinson, D., & Birmingham, P. (2003). *Using research instruments: A guide for researchers*. Psychology Press.

Willis, J. W., Jost, M., & Nilakanta, R. (2007). *Foundations of qualitative research: Interpretive and critical approaches*. Sage.

Wong, H. K., Britton, T., & Ganser, T. (2005). What the world can teach us about new teacher induction. *Phi Delta Kappan, 86*(5), 379–384. https://doi.org/10.1177/003172170508600509

Wubbels, T. (1992). Taking account of student teachers' preconceptions. *Teaching and Teacher Education*, *8*(2), 137–149. https://doi.org/10.1016/0742-051X(92)90004-M

Xu, H. (2017). Exploring novice EFL teachers' classroom assessment literacy development: A three-year longitudinal study. *The Asia-Pacific Education Researcher, 26*(3–4), 219–226. https://doi.org/10.1007/s40299-017-0342-5

Xu, Y., & Brown, G. T. (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. (2018). The impact of school-based assessment for qualifications on teachers' conceptions of assessment. *Assessment in Education: Principles, Policy & Practice, 25*(6), 638–654. https://doi.org/10.1080/0969594X.2017.1295020

Yazan, B. (2015). Three approaches to case study methods in education: Yin, Merriam, and Stake. *The Qualitative Report, 20*(2), 134–152.

Yin, R. K. (2002). *Case study research: design and methods*. Sage.

Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Sage.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education, 45*(4), 477–501. https://doi.org/10.1023/A:1023967026413

Zeng, W., Huang, F., Yu, L., & Chen, S. (2018). Towards a learning-oriented assessment to improve students' learning—a critical review of literature. *Educational Assessment, Evaluation and Accountability*, *30*(3), 211–250. https://doi.org/10.1007/s11092-018-9281-9

Zheng, H. (2013). The dynamic interactive relationship between Chinese secondary school EFL teachers' beliefs and practice. *The Language Learning Journal, 41*(2), 192–204. https://doi.org/10.1080/09571736.2013.790133

Zheng, H. (2015). *Teacher beliefs as a complex system: English language teachers in China*. Springer.

Zhu, Y., & Edwards, F. (2019). Teacher questioning in a Chinese context: Implications for New Zealand classrooms. *Teachers and Curriculum*, *19*(1), 27–33. https://hdl.handle.net/10289/13073

Appendices

Appendix A: The Letter for the programme leader in Phase 1

Epsom Campus
Gate 3, 74 Epsom Avenue
Auckland, New Zealand
Telephone 64 9 623 8899
Facsimile 64 9 623 8898
www.education.auckland.ac.nz
The University of Auckland
Private Bag 92601, Symonds Street
Auckland 1035, New Zealand

PARTICIPANT INFORMATION SHEET

(Programme leader)

Investigating pre-service Chinese-heritage language teachers' conceptions of assessment.

Researcher: Jiani Yun

My name is Jiani Yun, and I am a Doctor of Philosophy (PhD) student in the school of Curriculum and Pedagogy, Faculty of Education and Social Work, at the University of Auckland. I am currently undertaking a small-scale pilot study to investigate pre-service teachers' prior beliefs about assessment and how these beliefs may change due to participation in an initial teacher education programme. I would like to invite participants in your programme to be part of this research. The particular focus is on those who will go on to teach Chinese as an additional language in schools.

Project description and procedures

This project will investigate pre-service teachers' conceptions of assessment towards the end of a 1-year ITE programme at the University of Auckland. Only those who will go on to teach Chinese as an additional language in schools will be invited to participate. It is anticipated that teachers of Chinese in schools, particularly if they themselves come from a Confucian Heritage cultural background, will come into an initial teacher education programme with beliefs about assessment shaped by that cultural background that may be different to the thinking that informs most assessment in New Zealand's schools (in particular a difference between a summative knowledge based examination system and a formative learner-centred assessment system). This investigation into pre-service teachers'

conceptions of assessment will assist both pre-service teachers and teacher educators to understand pre-service teachers' underlying beliefs and help pre-service teachers to move towards establishing new beliefs that will promote effective assessment practice in schools.

I would like to conduct a focus group with up to eight pre-service teachers who are working towards becoming teachers of Chinese in schools in New Zealand. The aim of this project is to find out pre-service teachers' prior conceptions and beliefs about assessment, and the extent to which these change after engagement with a 1-year pre-service teacher education course.

If you agree to allow me to carry out this research in the programme, I will invite participants who fulfil the criteria to take part in one focus group interview which will be conducted in November 2016, just before participants are due to complete the initial teacher education programme.

The focus group will take up to 60 minutes, and it will be audiorecorded. The interview will be conducted at a convenient time for participants outside of lectures and tutorial time, and will most likely be held at the university campus. The researcher will transcribe the recording herself. There may be a need for short follow-up conversations to clarify certain terms or ideas, which may be by phone or email. A small thank-you gift will be offered for participants' support and participation.

Data storage, retention, destruction and future use

Digital data will be stored on the researcher's personal password protected computer. Hard copy data (such as transcripts) will be stored on the researcher's personal password protected computer or in a locked cabinet in her supervisor's office. Data will be stored for six years after completion of the research, and destroyed after this time. Data collected will only be used for academic and educational purposes, including as a small-scale pilot study for the researcher's PhD thesis at the University of Auckland, and potential academic publications and conference presentations.

Participants' rights

It is entirely up to participants to decide whether or not they wish to participate in this project. They are also free to withdraw this project before and during the focus group (without providing a reason), but not afterwards and not in relation to any contributions already made to the focus group. They are free not to answer any questions they do not wish to answer,

and they can ask for the recording to be turned off at any time. They will be informed that due to the nature of a focus group interview, the purpose of which is to gain insight into how participants represent a particular issue as a whole on a collective basis rather than an individual basis, participants will not be able to withdraw any comments they have made and will not be able to edit the transcript. If they would like to receive a copy of the transcript and a final summary of findings, they may indicate that on the consent form.

I would like to ask you to give your assurance that participation or non-participation will not affect participants' standing in the programme or their course grades in any way.

Anonymity and Confidentiality

Participants' identities and information will remain confidential to the researcher and her supervisor during this research. Anonymity of participants in the focus group is not possible. Participants will be asked to sign a consent form agreeing to keep the identities and comments of all members of the group confidential, and agreeing that what is discussed during the focus group will not be discussed outside the meeting. Due to the small number of participants, there is a small risk that participants may be identifiable in all final reports. No names will be used in any final reporting.

Thank you for considering allowing participants in your programme to participate in my research study.

CONTACT DETAILS AND APPROVAL

Student Researcher name and contact	Supervisor name and	Co-Supervisor name and	Head of Department name and contact
Jiani Yun School of Curriculum and Pedagogy Faculty of Education and Social Work The University of Auckland Email: j.yun@auckland.ac.nz	Associate Professor Martin East School of Curriculum and Pedagogy Faculty of Education and Social Work The University of Auckland Phone:(09) 623 8899 ext. 48345 Email: m.east@auckland.ac.nz	Associate Professor Mary Hill School of Teaching, Learning and Development Faculty of Education and Social Work University of Auckland Email: mf.hill@auckland.ac.nz	Associate Professor Helen Hedges Head of School School of Curriculum and Pedagogy Faculty of Education and Social Work The University of Auckland Phone: (9) 623 8899 ext. 48606 Email: h.hedges@auckland.ac.nz

For any queries regarding ethical concerns you may contact the Chair, The University of Auckland Human Participants Ethic Committee, The University of Auckland, Research Office, Private Bag 92019, Auckland 1142.

Telephone 09 373-7599 ext. 83711. Email: ro-ethics@auckland.ac.nz

Appendix B: The Consent Forms for the Programme Leader in phase 1

CONSENT FORM

(the programme leader)

This form will be held for a minimum period of 6 years

Project title: Investigating pre-service Chinese-heritage language teachers' conceptions of assessment

Researcher: Jiani Yun

I have read the Participant Information Sheet and have understood the nature of the research and why I have been selected. I have had the opportunity to ask questions and have them answered to my satisfaction. I understand that participation is voluntary.

I agree to participate in this research.

1. I understand that I will be given a pseudonym and that my privacy and confidentiality will be respected throughout the study and in any resulting reports or publications.
2. I have been provided an explanation of this research and understood its nature.
3. I have had an opportunity to ask questions and have them answered.
4. I give consent for the researcher to invite participants at this site and to conduct research as described in the participant information sheet.
5. I confirm that whether the Faculty member/s choose to participate or not, this will have no influence on their employment status.

Name _____   Signature: _____   Date: _____

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE ON 27 SEPTEMBER 2016 FOR THREE YEARS REFERENCE NUMBER 017964

Appendix C: Participant Information Sheets for Preservice Chinese-Heritage Language Teachers in Phase 1

Epsom Campus
Gate 3, 74 Epsom Avenue
Auckland, New Zealand
Telephone 64 9 623 8899
Facsimile 64 9 623 8898
www.education.auckland.ac.nz
The University of Auckland
Private Bag 92601, Symonds Street
Auckland 1035, New Zealand

PARTICIPANT INFORMATION SHEET

Research title: Investigating pre-service teachers' prior knowledge of assessment.

Researcher: Jiani Yun

My name is Jiani Yun, and I am a Doctor of Philosophy (PhD) student in the school of Pedagogy and Curriculum, Faculty of Education, the University of Auckland. I am currently undertaking research to investigate how teacher teacher s' prior knowledge of assessment are undertaking of assessment practicum. I sincerely invite you to participant in this research. The following information will give you an overview of this study, highlight information could be of reference to make the decision to participate effectively.

Project description and procedures

This project will investigate pre-service student teachers' conception of assessment toward the end of their 1-year initial teacher education program in the Auckland University. As shown in many research, student teachers' conceptions of assessment are critical components in their learning process, their prior knowledge at least cover 40%-50% of what teachers intend them to learn. Therefore, the investigation toward student teachers' prior knowledge of assessment could effectively assist teacher educator and student teachers themselves to scaffold their assessment knowledge system, promote their learning and practice.

I intended to conduct about qualitative case study of interview in teacher candidates' professional learning in University of Auckland and their assessment practice in their practicum schools. The aim of this project is to find out:

- What is pre-service teacher's prior knowledge and beliefs of assessment before they take assessment/evaluation relevant courses?

- What is pre-service teacher's conception and beliefs of assessment before they start their practicum?
- How pre-service teachers' prior knowledge and beliefs promote or hinder their assessment practice?

Your insights of assessment will be helpful to gain better understanding of prior knowledge of assessment and its impact on pre-service teachers' assessment learning and practice. If you agree to participant in this project, there will be three times of interviews: a) before you take assessment/evaluation course in semester Two; b) before you start your practicum in school; c) after you finish your practicum.

Each interview will take 30-40 minutes, and no more than 60 minutes. Interview (a), (b) will be conducted at a convenient for you outside of lectures and tutorial time, most likely at the university campus, interview (c) will be conducted at your school practicum outside of classroom time. There could be a need for follow-up conversations for clarify certain terms, which may be by phone or email.

You will be asked to share documents regarding to your learning about assessment and developing assessment capability in both University and practicum schools. For instance, assignments, your professional portfolio, lesson/unit plans, example of your students work.

If you agree to participate in my research, a 50-dollar gift card as thank-you gift will be offered my sincere gratitude for your support and participation.

Data storage, retention, destruction and future use
Data will be stored and protected either in the researchers' personal password protected computer or in her locked office cabinet. Data will be stored for six years after completion of the research, and deleted after this time, unless the researcher continues with research in this field. Data collected will only be used for academic and educational purposes, such as the researcher's PhD thesis at the University of Auckland, academic publications, conference presentations, teaching, and other forms of academic research dissemination.

Participants' rights
It's completely up to you to decide whether you will take part in this research. You can ask to receive the electronic copies of your interview and observation recordings and transcripts via email or cloud-delivery. You are able to change, delete and add further information on drafts of the transcripts.

Anonymity and Confidentiality

Your identity and information will remain confidential during this research and no personal features will be identified. You can choose a pseudonym in the attached Consent Form; otherwise the researcher will make up one for you.

Thank you for considering to participate in my research study.

Yours sincerely

Jiani Yun

If you wish to discuss any aspect of mu study with my supervisor Professor. Martin East and Professor. Mary Hill, you could contact:

Professor Martin East

School of Curriculum and Pedagogy

Faculty of Education

University of Auckland

Email: m.east@auckland.ac.nz

Dr. Mary Hill

School of Teacher Education Practice

Faculty of Education

University of Auckland

Email: mf.hill@auckland.ac.nz

For any queries regarding ethical concerns you may contact:

The Chair
The University of Auckland Human Participants Ethics Committee
The University of Auckland Office of the Vice Chancellor
Private Bag 92019
Auckland 1142
Telephone: (09) 373 7599 ext. 83711

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE ON 27 SEPTEMBER 2016 FOR THREE YEARS REFERENCE NUMBER 017964

Appendix D: The Consent Forms for Participants in Phase 1

CONSENT FORM (PRE-SERVICE TEACHERS)

(This consent form will be stored for a period of six years)

Research Study: Investigating pre-service teachers' conceptions of assessment

Researcher: Jiani Yun

I have read the Participant Information Sheet and have understood the nature of the research and why I have been selected. I have had the opportunity to ask questions and have them answered to my satisfaction. I understand that participation is voluntary and that I can withdraw from the study at any time up until 31 December 2016.

I agree to take part in this study. I understand that:

- I am being asked to partake in a focus group interview with other programme participants, and that the focus group will last up to 60 minutes

- Permission has been obtained from the Director of the Teacher Education programme for this study to occur and assurance has been given that my participation, or non-participation, will not affect my relationship with the university or my grade in any way.

- My privacy and confidentiality will be respected throughout the study and in any resulting reports or publications.

- The focus group will be digitally audiorecorded. I am free not to answer any questions I do not wish to, and I can ask for the recording to be turned off at any time.

- Due to the nature of the focus group, I will not be able to withdraw any information I have provided in the focus group.

- Data will be stored securely and separately from Consent Forms, in password protected computer and/or in locked cabinets at the Faculty of Education for six years, after which they will be destroyed.

I agree to maintain confidentiality of information shared in this focus group.

I give permission for information to be used as a pilot study for Jiani Yun's PhD thesis, possible publication in educational journals, possible presentations and as the basis for further study.

If I wish to receive a copy of the transcript / summary report I will give my email address below:

Full name: _____

Signed: _____ Email: _____

Date: _____

Phone: _____ Wechat: _____

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE ON 27 SEPTEMBER 2016 FOR THREE YEARS REFERENCE NUMBER 017964

Appendix E: Guiding Questions for the Focus Group Interview in Phase 1

Questions for the focus group interview

- What do you think the word 'assessment' means?

- What different types of assessments are you aware of? What do you think are the aims and purposes for different assessments?

- Please tell me more about your experience of assessment practice in China.

- Tell me about your own experiences of taking assessments

- How were your experiences different to what you saw on practicum or learnt about in the programme?

- What helped you most in your teacher education programme to learn about the purposes of assessment?

- Is there any difference between what you have learned about assessment through this programme and your previous learning/teaching experiences? Is so, please give examples.

- Based on your learning/teaching experiences, what do you think are the best methods to implement assessment in the classroom?

- Please tell me about the experiences, other than the assessment course, that assisted you to learn how to use assessment in the New Zealand context?

- What parts of assessment do you feel most confident about? What do you feel least confident to use? Why is this?

- Have you changed your knowledge/understandings/beliefs about assessment during this period of learning about assessment? Could you please give examples?

- What do you think might help you to learn more about assessment as you think about becoming a teacher?

Appendix F: Participant Information Sheets in Phase 2

Epsom Campus
Gate 3, 74 Epsom Avenue
Auckland, New Zealand
Telephone 64 9 623 8899
Facsimile 64 9 623 8898
www.education.auckland.ac.nz
The University of Auckland
Private Bag 92601, Symonds Street
Auckland 1035, New Zealand

PARTICIPANT INFORMATION SHEET

(BEGINNING TEACHERS)

Investigating beginning Chinese-heritage language teachers' conceptions and practice.

Researcher: Jiani Yun

My name is Jiani Yun, and I am a Doctor of Philosophy (PhD) student in the school of Curriculum and Pedagogy, Faculty of Education and Social Work, at the University of Auckland. I am currently undertaking a study to investigate beginning Chinese-heritage language teachers' conceptions and practices of assessment within first two years of working in New Zealand secondary schools. I would like to invite you to be part of this research. The particular focus is on those who have Chinese-heritage background and undertake language teaching in New Zealand secondary schools.

Project description and procedures

This project will explore beginning Chinese-heritage language teachers' conceptions and practices of assessment within two academic years (end 2017 and 2018) before they are fully registered. Only those who have Chinese-heritage backgrounds will be invited to participate. It is anticipated that novice teachers who come from a Confucian Heritage cultural background, may come to New Zealand secondary schools with beliefs about assessment shaped by Chinese Confucius heritage culture which may be different to the thinking that informs assessment in New Zealand schools (in particular a difference between a summative knowledge based examination system and a formative learner-centred assessment system). This investigation will assist both novice Chinese-heritage language teachers and teacher educators to gain a better understanding of novice Chinese-heritage teachers' underlying beliefs and help novice Chinese-heritage teachers to move towards establishing new beliefs that will promote effective assessment practice in schools.

I would like to conduct a 2-year case study with beginning Chinese-heritage language teachers who are working in New Zealand secondary schools. The aim of this project is to find out beginning Chinese-heritage language teachers' existing conceptions about assessment, and the extent to which these change within two years of assessment practices in different school cultures.

If you agree to participate in this project, I will conduct four times of classroom observations within academic year 2017 and 2018. The first one will be conducted at the end of academic year of 2017, the second will be at the beginning of academic year of 2018, the third will be in the middle and the last will be at the end of 2018. Reflective conversations to clarify certain terms or ideas will take up to 20 minutes after each classroom observation. Interviews will be conducted after the classroom observations, specifically these will take up to 60 minutes, at a convenient time and location for participants outside of their teaching commitments

Both classroom observations and interviews will be audiorecorded. As Mandarin and English may appear in classroom observations and interviews, I will transcribe and translate the recordings myself. The transcripts will be reviewed for accurateness of translation from Mandarin to English. Another PhD candidate who has signed the confidentiality agreement will review the transcripts.

If you agree to participate in my research, a thank-you gift of a $50 gift voucher will be offered for your support and participation.

Data storage, retention, destruction and future use

Digital data will be stored on my personal password protected computer. Hard copy data (such as transcripts) will be stored in a locked cabinet in my supervisor's office. Data will be stored for six years after completion of the research, and destroyed after this time. Data collected will only be used for academic and educational purposes, including my PhD thesis at the University of Auckland, and potential academic publications and conference presentations.

Participants' rights

It is entirely up to you to decide whether or not you wish to participate in this project. Participants are free to withdraw the data you have provided within a set period of two weeks after each data collection at four different points over two years without giving a reason. It will be impossible to withdraw data after that time because the analysis of

research results will be underway. I will transcribe and translate the recordings, and the participants will have the option to edit the transcripts. If you would like to receive a copy of the transcript and a final summary of findings, you may indicate that on the consent form.

Please be aware that the school leader has given his/her consent for the research to take place in the school, and has given an assurance that your participation or non-participation will not affect your standing in the school in any way.

Anonymity and Confidentiality

The participant's identities will remain confidential to me and my supervisors during this research. Pseudonyms will be used throughout the whole study.

Thank you for considering participation in my research study.

CONTACT DETAILS AND APPROVAL

Student Researcher name and contact	Supervisor name and contact	Co-Supervisor name and contact	Head of Department name and contact
Jiani Yun School of Curriculum and Pedagogy Faculty of Education and Social Work The University of Auckland Email: j.yun@auckland.ac.nz	Associate Professor Martin East School of Curriculum and Pedagogy Faculty of Education and Social Work The University of Auckland Phone:(09) 623 8899 m.east@auckland.ac.nz	Associate Professor Mary Hill School of Teaching, Learning and Development Faculty of Education and Social Work University of Auckland Email: mf.hill@auckland.ac.nz	Associate Professor Helen Hedges Head of School School of Curriculum and Pedagogy Faculty of Education and Social Work The University of Auckland Phone: (9) 623 8899 h.hedges@auckland.ac.nz

For any queries regarding ethical concerns you may contact the Chair, The University of Auckland Human Participants Ethic Committee, The University of Auckland, Research Office, Private Bag 92019, Auckland 1142.

Telephone 09 373-7599 ext. 83711. Email: ro-ethics@auckland.ac.nz

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE ON 11 NOVEMBER 2017 FOR THREE YEARS REFERENCE NUMBER 020383

Appendix G: Consent Form in Phase 2

CONSENT FORM

(BEGINNING TEACHERS)

(This consent form will be stored for a period of six years)

Investigating beginning Chinese-heritage language teachers' conceptions and practice.

Researcher: Jiani Yun

I have read the Participant Information Sheet and have understood the nature of the research and why participants have been selected. I have had the opportunity to ask questions and have them answered to my satisfaction. I understand that participation is voluntary and that participants can withdraw from the study at any time up until 31 December 2018.

I agree to take part in this study. I understand that:

- Classroom observations will be conducted four times within the academic years of 2017 and 2018. The first observation will be conducted at the end of the academic year 2017, the second will be in the beginning of academic year 2018, the third will be in the middle and the last will be at the end of academic year of 2018.

- Reflective conversations to clarify certain terms or ideas will take up to 20 minutes after each classroom observation.

- Interviews will last up to 60 minutes and will take place outside the school hours.

- Participants' privacy and confidentiality will be respected throughout the study and in any resulting reports or publications.

- The classroom observations and interviews will be digitally recorded. Participants are free not to answer any questions they do not wish to answer, and they can ask for the recording to be turned off at any time.

- Another PhD candidate who has signed the confidentiality agreement will review the transcripts.

- Data will be stored securely in a locked cabinet at the Faculty of Education and Social work for six years, after which they will be destroyed.

- Data will be used as a main study for Jiani Yun's PhD thesis, possible publication in educational journals, and possible presentations.

I give permission for information to be used for Jiani Yun's PhD thesis, possible publication in educational journals, and possible presentations.

If I wish to receive a copy of the transcript / summary report I will give my email address below:

Full name: _____

Signed:_____Email:_____

Date: Phone:_____

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE ON 11 NOVEMBER 2017 FOR THREE YEARS REFERENCE NUMBER 020383

Appendix H: The Example of Collected Documentary Data (Diana's Course Outline and Assessment Schedules)

Year 10 Chinese 2018
Course Outline and Assessment Schedule

Term	Week	Date	Topic	Assessment
			TERM 1	
Term 1	1	22-Jan	Introduction and Commendation	
	2	29-Jan	Introduction and Commendation	
	3	5-Feb	Introduction and Commendation	
	4	12-Feb	Introduction and Commendation/Chinese New Year	
	5	19-Feb	Introduction and Commendation	
	6	26-Feb	Introduction and Commendation	Progress Test 1 (L/R/W, 30%)
	7	5-Mar	Introduction and Commendation	
	8	12-Mar	City and Environment	
	9	19-Mar	City and Environment	
	10	26-Mar	City and Environment	
	11	2-Apr	City and Environment	Progress Test 2 (L/R/W, 30%)
	12	9-Apr	City and Environment	Cultural Project 1 (10%)
		16-Apr	Study Break	
		23-Apr		
			TERM 2	
Term 2	1	30-Apr	City and Environment	
	2	7-May	City and Environment	
	3	14-May	EXAMINATIONS (Listening & Speaking)	
	4	21-May	EXAMINATIONS (Reading & Writing)	
	5	28-May	House and Furniture	
	6	4-Jun	House and Furniture	
	7	11-Jun	House and Furniture	
	8	18-Jun	House and Furniture/ Dragon Boat Festival	
	9	25-Jun	House and Furniture	Cultural Project 2 (10%)
		2-Jul	Study Break	
		9-Jul		
		16-Jul		
			TERM 3	
Term 3	1	23-Jul	House and Furniture	
	2	30-Jul	House and Furniture	Progress Test 3 (L/R/W, 30%)
	3	6-Aug	School Life	
	4	13-Aug	School Life	
	5	20-Aug	School Life	
	6	27-Aug	School Life	
	7	3-Sep	School Life	
	8	10-Sep	School Life	
	9	17-Sep	Health	Progress Test 4 (L/R/W, 30%)
	10	24-Sep	Health/Full Moon Festival	
		1-Oct	Study Break	
		8-Oct		
			TERM 4	
Term 4	1	15-Oct	Health	
	2	22-Oct	Health	
	3	29-Oct	EXAMINATIONS (Listening & Speaking)	
	4	5-Nov	EXAMINATIONS (Reading & Writing)	
	5	12-Nov	Health	
	6	19-Nov	Health	
	7	26-Nov	Chinese Cuisine	
	8	3-Dec	Chinese Cuisine	

Appendix I: The Example of Collected Documentary Data (Amber's Notes of the Speaking Test)

	QUESTIONS	STUDENT ANSWERING	STUDENT ASKING
1	你好吗？	✓	✓
2	你叫什么名字？		
3	你姓什么？	是 ✓	✓
4	你的电话号码是多少？		
5	你几岁？		✓
6	今天是星期几？		
7	你的生日是几月几号？		✓
8	你属什么？		我属什么 ✗ self-corrected
9	你会不会说英文？		
10	你是哪国人？	✓	
11	你是不是中国人？		
12	你有没有兄弟姐妹？		
13	你家有几口人？	✓	
14	你家有什么宠物？		
15	你家有没有宠物？	✓	
16	你家有几个房间（客厅/卫生间）？	✓	
17	你喜欢不喜欢（name of leisure activity）？	✓	
18	你喜欢不喜欢（name of sports）？		
19	你喜欢什么运动？		
20	你的爱好是什么？	✓	
21	你们班有几个学生？		
22	你在哪个班？	✓	
23	你上几年级？		
24	你上哪个学校？		
25	你星期一有什么课？	✓	

Appendix J: Sample Codes From Themes That Resulted From Data Analysis

Appendix K: The Example of Analysed Documentary Data (Michelle's Exam Notes)

PARTICIPANT INFORMATION SHEET

(School Leader)

Investigating beginning Chinese-heritage language teachers' conceptions and practice.

Researcher: Jiani Yun

My name is Jiani Yun, and I am a Doctor of Philosophy (PhD) student in the school of Curriculum and Pedagogy, Faculty of Education and Social Work, at the University of Auckland. I am currently undertaking a study to investigate novice Chinese-heritage teachers' conceptions and practices of assessment within first two years of working in New Zealand secondary schools. I would like to invite participants in your school to be part of this research. The particular focus is on those who have Chinese-heritage background and undertake language teaching in New Zealand secondary schools.

Project description and procedures

This project will explore novice Chinese-heritage teachers' conceptions and practices of assessment within two academic years before they are fully registered. Only those who have Chinese-heritage backgrounds will be invited to participate. It is anticipated that novice teachers who come from a Confucian Heritage cultural background, will conduct assessment tasks in New Zealand secondary schools with beliefs about assessment shaped by Chinese Confucius heritage culture which may be different to the thinking that informs most assessment in New Zealand schools (in particular a difference between a summative knowledge based examination system and a formative learner-centred assessment system). This investigation into novice Chinese-heritage teachers' conceptions of assessment will assist both novice teachers and teacher educators to understand novice Chinese-heritage

teachers' underlying beliefs and help novice Chinese-heritage teachers to move towards establishing new beliefs that will promote effective assessment practice in schools.

I would like to conduct a 2-year longitudinal case study with four novice Chinese-heritage teachers who are working in New Zealand secondary schools. The aim of this project is to find out novice Chinese-heritage teachers' existing conceptions about assessment, and the extent to which these change within two years of assessment practices in different school cultures.

If you agree to allow me to carry out this research in your school, I will invite the participant who fulfils the criteria to take part in my research. I will conduct four times of classroom observations within academic year 2017 and 2018. One will be conducted at the end of academic year of 2017, one will be at the beginning of academic year of 2018, one will be in the middle and one will be at the end of 2018. The interview will be conducted within two days after the classroom observation, specifically it will take up to 60 minutes, at a convenient time for participants outside of lectures and tutorial time, and will most likely be held in the community libraries nearby

Both classroom observations and interviews will be audiorecorded. The researcher will transcribe the recording herself. Short reflective conversations to clarify certain terms or ideas will be conducted after each classroom observation. A thank-you gift will be offered for participants' support and participation.

Data storage, retention, destruction and future use

Digital data will be stored on the researcher's personal password protected computer. Hard copy data (such as transcripts) will be stored on the researcher's personal password protected computer or in a locked cabinet in her supervisor's office. Data will be stored for six years after completion of the research, and destroyed after this time. Data collected will only be used for academic and educational purposes, including the researcher's PhD thesis at the University of Auckland, and potential academic publications and conference presentations.

Participants' rights

It is entirely up to participants to decide whether or not they wish to participate in this project. As for classroom observation data, it will be impossible or impractical for the researcher to remove the data once it has been collected. The reason is that this data is generated through

interactions in the classroom, and the removal may affect any data generated in response to the withdraw data. As for interview data, they are entitled to withdraw the data they have provided within a set period of two weeks after data collection, without giving a reason. It will be impossible to withdraw data after that time because the analysis of research results will be underway. The researcher will transcribe the recordings, and the participants will have the option to edit the transcripts and to request a copy of the recordings. If they would like to receive a copy of the transcript and a final summary of findings, they may indicate that on the consent form.

I would like to ask you to give your assurance that participation or non-participation will not affect participants' standing in the school in any way.

Anonymity and Confidentiality

Participants' identities and information will remain confidential to the researcher and her supervisor during this research. Pseudonyms will be used throughout the whole study. The confidentiality of the data will be preserved throughout the time it is required to be stored.

Thank you for considering allowing participants in your school to participate in my research study.

CONTACT DETAILS AND APPROVAL

Student Researcher name and	Supervisor name and contact	Co-Supervisor name and	Head of Department name and contact
Jiani Yun School of Curriculum and Pedagogy Faculty of Education and Social Work The University of Auckland Email: j.yun@auckland.ac.nz	Associate Professor Martin East School of Curriculum and Pedagogy Faculty of Education and Social Work The University of Auckland Phone:(09) 623 8899 m.east@auckland.ac.nz	Associate Professor Mary Hill School of Teaching, Learning and Development Faculty of Education and Social Work University of Auckland Email: mf.hill@auckland.ac.nz	Associate Professor Helen Hedges Head of School School of Curriculum and Pedagogy Faculty of Education and Social Work The University of Auckland Phone: (9) 623 8899 h.hedges@auckland.ac.nz

For any queries regarding ethical concerns you may contact the Chair, The University of Auckland Human Participants Ethic Committee, The University of Auckland, Research Office, Private Bag 92019, Auckland 1142.

Telephone 09 373-7599 ext. 83711. Email: ro-ethics@auckland.ac.nz

Appendix M: Consent Form for Secondary School Leaders

Epsom Campus
Gate 3, 74 Epsom Avenue
Auckland, New Zealand
Telephone 64 9 623 8899
Facsimile 64 9 623 8898
www.education.auckland.ac.nz
The University of Auckland
Private Bag 92601, Symonds Street
Auckland 1035, New Zealand

CONSENT FORM (SCHOOL LEADER)

(This consent form will be stored for a
period of six years)

**Investigating beginning Chinese-heritage language
teachers' conceptions and practice.**

Researcher: Jiani Yun

I have read the Participant Information Sheet and have understood the nature of the research and why participants have been selected. I have had the opportunity to ask questions and have them answered to my satisfaction. I understand that participation is voluntary and that participants can withdraw from the study at any time up until 31 December 2018.

I agree that participants in the school may take part in this study.

I understand that:

- Classroom observations will be conducted four times within the academic years of 2017 and 2018. The first observation will be conducted at the end of the academic year 2017, the second will be in the beginning of academic year 2018, the third will be in the middle and the last will be at the end of academic year of 2018.

- Reflective conversations to clarify certain terms or ideas will take up to 20 minutes after each classroom observation.

- Interviews will last up to 60 minutes and will take place outside the school hours.

- Participants' privacy and confidentiality will be respected throughout the study and in any resulting reports or publications.

- The classroom observations and interviews will be digitally recorded. Participants are free not to answer any questions they do not wish to answer, and they can ask for the recording to be turned off at any time.

- Another PhD candidate who has signed the confidentiality agreement will review the transcripts.

- Data will be stored securely in a locked cabinet at the Faculty of Education and Social work for six years, after which they will be destroyed.

- Data will be used as a main study for Jiani Yun's PhD thesis, possible publication in educational journals, and possible presentation

I agree to allow for this study to occur in my school and give my assurance that participation, or non-participation, will not affect participants' relationships with the school in any way.

If I wish to receive a copy of the transcript / summary report I will give my email address below:

Full name: _____

Signed:_____ Email: _____

Date: _Phone: _____