Assessing for learning in middle school English language classrooms in China

Jingwei Song

BA (English language literature)

MA (Teaching of English to Speakers of Other Languages)

Submitted in fulfilment of the requirements of the

Degree of Doctor of Philosophy (PhD)

School of Education, College of Social Sciences

University of Glasgow

May 2023

Abstract

As a growing economy and a fast-changing society, China has recognised the important role that education and the learner journey must play in that transformation. Assessment is a significant part of a student's learning journey. It has the potential to engage students in the development of clear learning goals, to engage them in reflecting on their progress and performance, and to provide a basis for constructive feedback on how further progress can be made. In China, educational reforms featuring policies advocating formative assessment have sought to improve assessment practices in English language classrooms. Previous studies have explored teachers' assessment practices and understandings in tertiary English education. However, assessment in Chinese secondary English classrooms, particularly those in middle schools, receives little attention. Local responses to the national assessment policies also receive inadequate investigations. This study addresses these issues.

The study adopts a multiple-case study approach and investigates four English teachers and their classes in two middle schools in Shenzhen, China, in the context of the new 'Zhongkao' (senior high school entrance examination) reform in Shenzhen. Three research methods are chosen and findings triangulated. First, materials including textbook, teacher guidebook, and English Zhongkao exam paper are analysed using content analysis approach to understand the assessment content teachers work with. Second, each teacher's assessment activities are explored through classroom observations over a unit of teaching. Third, teachers' understanding of assessment and assessment policies is investigated through before- and after-observation interviews and analysed using thematic analysis approach.

The data analysis reveals that the teachers adopt three types of assessment activities – oral assessment activities, written assessment activities, and student-assessed activities – with oral assessment activities being conducted the most frequently and student-assessed activities the least often. The teachers implement these assessment activities for various purposes, including assessing for instruction, learning, and maintaining discipline. Analysis of the assessment context demonstrates a wide range of factors inside and outside of classrooms influencing the teachers' assessment activities and understandings. These include, first, teachers' limited past academic and professional education regarding assessment, which poses a barrier for them in carrying out formative assessment practices; second, stakeholders' test-result-oriented expectations, which provide a basis for teachers' test-oriented aspirations for the future; third, teachers' working environment, which has exerted an impact on teachers' actions from three levels: the classroom level involves class size and students' language level, the school level involves accountability pressure and available assessment support, and the policy level involves the Zhongkao and mandated textbook.

The significance of the study is threefold. First, it contributes to the understanding of Chinese middle school English teachers' assessment activities and understanding. Second, it has rich implications for the Zhongkao reform in Shenzhen regarding test design and washback on teachers. Third, it proposes a framework for understanding classroom assessment activities and teachers' assessment beliefs from a contextual perspective, which may be adopted and adapted for assessment research in other contexts.

Contents

List of tables

List of figures

Acknowledgment

First and foremost, I would like to acknowledge the financial support from China Scholarship Council and University of Glasgow. Without the scholarship, conference support, and other financial aids, completing this study on time during the COVID pandemic would be difficult.

Second, I would like to thank my supervisors, Professor Louise Hayward, Dr Sally Zacharias, and Professor Emerita Catherine Doherty. They have led me into the academic world with their expertise, patience, kindness, and lovely humour. It is an honour and privilege to work with you all. I would not have made it so far without your support. I would also like to thank Professor Qiuxian Chen for the valuable inspiration I achieved during our conversations.

Thirdly, I would like to express my gratitude to the participating schools, teachers, and students in this study. I would like to thank Ms. Kuanping Xie and Ms. Yuemei Ning for their generous assistance in gaining access to schools. I would also like to pay my tribute to all the medical workers, community volunteers, and all other people who fight COVID in their own way. Without their efforts, my visit to schools and completing this study would be impossible.

Fourthly, my gratitude goes to my family, who brought me up and supported my education. I would like to thank my mom for being a great listener and for her understanding and encouragement. I am also grateful for having many wonderful friends, who listened to my complaints, offered a hand to me when I encountered difficulties in life and study, and encouraged me to look on the bright side.

Finally, I want to thank my partner Xipeng Liu for inspiring me to pursue a PhD and for accompanying me during my PhD study as a friend, a colleague, and a great chef. Not many people do their PhD with their favourite person in the world doing the same thing – you make this amazing thing happen to me.

Author’s declaration

I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.”

Printed Name: _____Jingwei Song_____

Signature: _____

Chapter 1 Introduction

Having been an important part of the history of education in China, assessment has been historically associated with social status, occupational future, and success (Han & Yang, 2001). Such a history profoundly impacts educational practices in China nowadays and has constant collisions with emerging educational beliefs. In this chapter, I will first briefly introduce the historical background to the study, including the history of assessment reform in China (section 1.1), which dates back to imperial China, and the history of the Chinese educational context and English curriculum reform since 1949, when the People's Republic of China (PRC) was founded (section 1.2). I will then introduce the current curriculum and assessment policies since 2000 (section 1.3) to illustrate the current background of the study. Based on the history and the current situation, I will discuss the research problems that need to be investigated by this study (section 1.4), explain my motivation (section 1.5), and raise the research questions that need to be answered (section 1.6).

1.1 History of assessment reform in China

China has a long history of assessment practices dating back to the Western Zhou Dynasty (1027–771 BC) (Han & Yang, 2001). In ancient imperial China, competitive examinations dominated the traditional education system and played an essential role in selecting government officials (Berry, 2011b; Cheng & Curtis, 2010). One of the most important assessment systems in ancient China is called 'Keju' (科举制度), the imperial examination system, which started in the Sui Dynasty (around 607 AD) and was abolished in 1905 when the Qing Dynasty was facing its end (Fan, 2006). The system focused on assessing candidates' scholastic achievements through formal examinations (Berry, 2011b). It required the candidates to recite the Confucian classics and write essays to discuss a specific statement made in the Confucian classics (Wang, 2016). Under this system, examinations have been considered a general approach to encourage academic development and, to some extent, an opportunity to oppose corruption and maintain equity (Cheng & Curtis, 2010). While discussions have been held on the negative impact of the Keju system, such as narrowing academic achievement to Confucian classics and making people believe that the Keju test was the only purpose for learning (Fan, 2006), it is crucial to recognise that the system was an advanced mechanism in a feudal society and has its own contribution to the socio-economic development at its time.

Changes were made to the assessment system in the late nineteenth century before the Qing Dynasty collapsed. With the increase in interactions between the East and the West, social reforms such as 'Westernisation Movements' (洋务运动) and "Hundred Days' Reform" (戊戌变法) were triggered by the reform-minded elites and provided new directions for modern educational assessment policies (Berry, 2011b). According to Berry (2011b, p. 50), after being abolished in 1905, Keju was replaced by "a three-tier national examination system", which targeted "assessing students at the end of the three major stages of schooling – primary, middle and senior secondary". In addition, various types of examinations were introduced and developed to meet the needs in different periods of schooling, for example, entry exams, mid-term exams, non-regular tests, and final exams (Berry, 2011b; Han & Yang, 2001). Although massive changes were made in the assessment system's structure, the assessment function remained unchanged in this period, as examinations were still used to make summative judgements and measure success (Berry, 2011b). Development in education at this stage was slow because of the ongoing wars from the 1920s to the 1940s. It was not until after the founding of the PRC that there were further changes in the assessment system.

After the founding of the PRC in 1949, the Chinese government adopted a policy of "combining mass education with elite education" to reform the education system (Han & Yang, 2001, p. 6). This policy had two goals: increasing access to basic education and training administrative officials and specialised talents in higher education institutions. 'Gaokao' (高考), the national college entrance examination, was launched in 1952 to ensure that academically-talented people had access to the limited higher educational resources (Song, 2005). Foreign languages were among the tested subjects, and candidates could choose to take an English test or a Russian test (Zheng & Xu, 2020). At this stage, some attempts to develop teacher-based assessment were made, which focused more on the process of students' learning. For example, according to Berry (2011b, p. 51) and Han and Yang (2001, p. 7), China borrowed "a five-grade-marking system" from the Soviet Union to replace percentiles with gradings for "judgments of performances (Distinction 5, Good 4, Pass 3, Fail 2, Poor 1) in schools". Students were graded according to their understanding and application of knowledge and written and verbal presentation skills, but whether this was a practice that contributed to students' learning process is yet to be explored. However, as Han and Yang (2001) reported, the attempts did not change the fact that more attention was given to examinations. The learning burden on students increased, and the curriculum was narrowed because "teachers emphasised teaching based on examination-oriented

education", for teacher evaluation was linked to students' exam results (Han & Yang, 2001, p. 7).

Following the tension between China and the Soviet Union and the Cultural Revolution between the 1960s and 1970s, the newly implemented assessment practices were halted. The Gaokao was abolished and replaced by a recommendation mechanism, as people at that time considered that Gaokao was not as fair as it claimed and that higher educational resources should be opened to the proletariats regardless of their academic achievement (Song, 2005). This act resulted in a sharp decline in the quality of education (Shi, 2001). The situation only improved after the Gaokao was restored in 1977 when the Reform and Opening-up Policy (改革开放) was adopted. Academic curricula in schools were back on track, and respect for knowledge was stimulated (Fan, 2006). Zhongkao (中考), the senior high school entrance exam, was also piloted in several provinces and metropolitans in the 1980s and was later implemented in the country to select students from middle schools to academic high schools or vocational high schools (Vickers & Zeng, 2017). English as a subject was brought into the education and assessment system due to China's need to open its market and re-establish international relationships, which triggered a fast development in English language education (Cheng & Curtis, 2010; Wang, 2007).

The public applauded the restoration of Gaokao by then because it met "the traditional expectations of fair distribution of opportunities and the expectation of a wise, powerful, central leadership that would prevent social injustice" (Feng, 1999, p. 44). However, exam-oriented education was also restored and strengthened simultaneously because the competition to enter university was increasing (Han & Yang, 2001). Such a phenomenon caused problems at different levels of education and aroused controversy. One of the most obvious problems was that test scores were closely associated with the evaluation of teachers and students (Song, 2005). According to Han and Yang (2001), the promotion rates of students moving up to key institutes were crucial for schools in some areas because educational authorities and parents would use these figures to judge the educational quality of schools. Consequently, teachers were driven to focus on helping students pass the selective exams, and students were trained particularly on taking the exams. This cast students in a passive role and gave them limited training in solving real-life problems (Ministry of Education, 1999).

To tackle this issue, the Ministry of Education instituted another round of assessment reform in the late 1990s, and attempts to promote formative assessment were made to create a better balance in the assessment system (Berry, 2011b; Han & Yang, 2001). New policies, guidelines, and instructions for assessment practices were provided, many of which were integrated into the curriculum documents released later in the 2000s. Along with the launch of the 'quality education' (素质教育) strategy, the government sent out strong signals that the assessment system should not only be used for selection purposes but also for enhancing teaching and supporting learning (Berry, 2011b). The Ministry of Education, therefore, issued a document called 'Announcement to Actively Promote the Reform of Evaluation and Examination Systems at Primary and Secondary Levels' (关于积极推进中小学评价与考试制度改革的通知) in 2002, requiring all schools to implement both formative and summative assessment in their educational planning. The role of assessment was also redefined as the key to "promoting quality-based education" and "enhancing the all-round development of students" (Han & Yang, 2001, p. 8). Although great effort has been made to promote assessment reform, challenges still exist. Wang (2008) reported two significant inadequacies in implementing formative assessment: first, many teachers were unprepared for the new assessment concept; second, many teachers did not know how to integrate assessment into teaching and learning. Berry (2011b) also suggested that no concrete suggestions were available to instruct teachers in classroom assessment designs.

Overall, China has been through a long journey in assessment reform. From the ancient 'Keju' to 'Gaokao', then to the current promotion of formative assessment, the definitions and functions of assessment have changed according to the educational purposes and the needs of different periods of society. The assessment history in China shapes how people perceive assessment and assessment policies. However, how teachers reflect on the current assessment policies and how these reflections affect their assessment practices is yet unknown. Among the major educational changes over time, curriculum reforms often took place alongside assessment reforms. Both reforms represented changes in the nation's curricular objectives and educational philosophy. In the following section, a historical overview of the educational context in China will be presented. The changes in the English curriculum in each historical period will be discussed, and the shift of underlying educational beliefs will also be outlined.

1.2 History of educational context and English curriculum reform in China since 1949

Since the founding of the PRC in 1949, education policies in China have changed almost every decade for socio-political and economic reasons. Foreign language teaching in basic education has also been adjusted from time to time to meet the emerging needs of the nation (Gu, 2012). Shi (2001) suggests six periods in which curriculum policies have changed and developed in China: the Soviet period, the exploring period, the Cultural Revolution period, the restoration period, the compulsory education period, and the quality education period. The foreign language curriculum has also developed throughout these periods, as discussed below.

In 1949, at the start of the Soviet period, the PRC was built on the ruins of several years of war. Although establishing an education system was among the priorities of nation-building, there was not enough time for the government to prepare a new curriculum that could fulfil the needs of the new nation (Fu, 1986). As a result, the government turned to the Soviet Union, its most powerful ally at that time, for experience and expertise in education policy planning. The curriculum policies issued in the 1950s were firmly based on Soviet education policies, and the syllabi and textbooks of all subjects in basic education reflected the textbooks used in Soviet schools (Shi, 2001). Regarding foreign language education, from 1949 to the late 1950s, Russian was the dominant foreign language taught in school because of the close relationship between China and the Soviet Union. English was given less attention in basic and tertiary education (Fu, 1986).

The exploring period began after China gradually distanced from the Soviet Union in the 1960s. Problems emerged as the borrowed curriculum was divorced from the Chinese reality and lacked coherence and quality (Gu, 2012). The government realised the importance of establishing a coherent curriculum that suited the Chinese context, and a short period of curricular and textbook exploration and experimentation emerged (Shi, 2001). In this period, the Ministry of Education was responsible for formulating national standards of curriculum and syllabus, and local authorities could design and publish local textbooks. By 1965, the government had made remarkable progress in implementing a curriculum that aimed at “two fundamentals”: knowledge and skills (Gu, 2012, p. 36). During this period, Russian fell out

of favour, and English replaced Russian as the major foreign language taught in schools due to the nation's need for English-speaking talents (Fu, 1986).

From 1966 to 1976, the Cultural Revolution, a period of social, political, and economic chaos in China, led to a disruption in the curriculum reform (Cheng & Curtis, 2010). Constant political assemblages disrupted education at all levels; everything foreign was rejected, including teaching and learning foreign languages such as English (Fu, 1986). Turning points in this period appeared when the PRC replaced the Republic of China in the United Nations in 1971 and when Richard Nixon, the U.S. president, visited China in 1972. These political events on the international stage raised efforts to cultivate more translating and interpreting personnel and re-establish English as a subject in schools (Gu, 2012). However, no English language syllabus was released in this period, and the chaos in education practice came to an end only after the political power shift in 1976, which ended the Cultural Revolution.

Following the Chinese economic reform in 1978, education in the restoration period became one of the priorities in China's opening-up and modernisation, and the education system was restored to prepare a new generation for a new era (Gu, 2012). Some of the policies during the exploring period were reinstated. For example, the Ministry of Education oversaw the formalisation of the education system, and textbooks were drafted based on the new syllabus. The local authorities could also design textbooks to fit their local contexts, but the Ministry of Education retained the power to review the quality of the textbooks (Shi, 2001). In this restoration period, English became firmly embedded in the basic education system and became one of the main subjects taught in almost every school, along with Chinese and mathematics (Gu, 2012; Liu, 2008). The economic changes in terms of the opening-up and modernisation prepared a climate for this round of educational reform, which transformed the educational system into a more open and flexible one (Guan & Meng, 2007). There were two syllabi published during this period: the trial draft version of the 'English Language Syllabus for Full-time Ten-year Primary and Secondary Schools' (全日制十年制中小学英语教学大纲[试行草案]) in 1978 and a revised version issued in 1980. According to Gu (2012, p. 37), the two syllabi represented the start of a "utilitarian era" because they both emphasised "the traditional emphasis on pronunciation, grammar, and vocabulary, and on integrated instruction in listening, speaking, reading, and writing".

In 1986, the National People's Congress passed the 'Compulsory Education Law of the People's Republic of China' (中华人民共和国义务教育法), which placed compulsory education (nine years of primary and middle school education) on a firm legal basis in China and marked the beginning of the compulsory education period (Shi, 2001). Accordingly, a new round of curriculum reform was introduced. There were two major features of this period. Firstly, separate curriculum documents were issued for compulsory education and high school education (Gu, 2012; Shi, 2001). The 'English Language Syllabus for Nine-year Compulsory Education Full-time Primary Schools and Middle Schools (Preliminary Version)' (九年义务教育全日制小学、初级中学英语课程计划[试行]) was published in 1988, and the 'Full-Time Syllabus for High Schools (Experimental Version)' (全日制普通高级中学课程计划[试验修订稿]) was issued in 1996. The different needs in the two educational stages made it necessary to have separate instructional documents (Guan & Meng, 2007). Secondly, a three-level educational management model was implemented (Gu, 2012; Guan & Meng, 2007): the Ministry of Education was responsible for giving overall guidance and instructions; local governments developed their own curriculum and textbooks based on the local needs; schools decided what compulsory and elective courses they offered. The model intended to tackle the imbalance of educational resources and economic status between different regions and made it possible for local governments and schools to maintain flexibility and autonomy (Shi, 2001).

The new syllabi in this period incorporated several innovative features – for example, a more communicative pedagogy was encouraged, and non-linguistic features such as studies of foreign cultures, critical thinking ability, and learner autonomy were introduced (Adamson, 2001; Liu, 2008). These changes resulted from a survey conducted by the Ministry of Education in 15 provinces and metropolitans in 1986. The findings indicated that most high school graduates could not communicate with others in simple English, even after attending 900 hours of English lessons (Adamson, 2001; Wang, 2007). Such findings were interpreted as a consequence of grammar-translation and behaviourist teaching methods, a lack of qualified teachers, and rigid examinations. Therefore, the Ministry of Education decided that a more communicative approach should be included in the new educational reform (Adamson, 2001). It was a controversial decision, as some educationalists advocated that the communicative approach did not match the learning styles of Chinese students and the traditional culture in Chinese classrooms (Anderson, 1993). In response to the debate, the

Ministry of Education made a concession in the new curriculum and took a cautious approach to pedagogic reform by blending 'Western' practices with educational beliefs popular in China (Adamson, 2001; Wang, 2007). Overall, this was a move in the philosophy of curriculum design towards communicative language teaching.

In the early 2000s, China started its new national curriculum reform in basic education, and 'quality education' became a national education strategy, which has been held to date. Quality education was understood to focus on the overall development of children (Gu, 2012). It responded to China's fast-growing economy, rapid social changes, information technology development, and economic globalisation (Wang & Chen, 2012). It also demonstrated that the educational goal of China was no longer the selection of elite intellectuals but rather the cultivation of "globally-compatible and future-ready" students, which led to the need for further curriculum restructuring (Gu, 2012, p. 8). The Ministry of Education issued the 'Framework for the Curriculum Reform of Basic Education (Trial Version)' (基础教育课程改革纲要[试行]) in 2001, along with new curriculum standards for 22 subjects. These included two English curriculum documents: 'English Curriculum Standards for Full-time Compulsory Education and Senior High Schools (Trial Version)' (全日制义务教育、普通高级中学英语课程标准[实验稿]) published in 2001 and 'English Curriculum Standards for Senior High Schools (Trial Version)' (普通高中英语课程标准[实验]) released in 2003. In July 2010, the Ministry of Education issued the 'National Framework for Mid- to Long- term Educational Reform and Development (2010-2020)' (国家中长期教育改革和发展规划纲要[2010-2020 年]), which vowed to develop quality education further and provide equal educational opportunities for every child. In response to the 'Framework', the Ministry of Education released 'English Curriculum Standards for Compulsory Education' (义务教育英语课程标准), an updated version of the 2001 curriculum document, in 2011 and 'English Curriculum Standards for Senior High Schools' (普通高中英语课程标准), an updated version of the 2003 curriculum, in 2017.

To sum up, China's education policies and foreign language curricula have undergone six distinct historical periods over the last few decades. Each period has its features that result from emerging socio-political and economic reasons, and the foreign language curriculum changed over time according to the society's needs. In the shift from Russian to English, the language education policies reflected the nation's plan to maintain diplomatic relations and

embrace a more open stance and modernisation. The widespread use of English worldwide and the unique socio-cultural, educational, and linguistic contexts in China also contributed to the development of the English curricula. Throughout the educational reforms since 1949, curriculum and assessment have been closely combined, especially in English language education. Curriculum standards inform teachers about what to teach and how to teach. On the other hand, assessment affects the instructions in actual classroom settings. However, how the changing policies influence and reshape learning, teaching, and assessment in classrooms remains to be explored. In the next section, changes made in the English curriculum and assessment policies in the current 'quality education' period will be presented in more detail.

1.3 Curriculum and assessment policies of English education in the 'quality education' period since 2000

The underpinning belief of curriculum reform in the 'quality education' period stemmed from a prevailing assumption that the exam-oriented educational practices were creating "high scores and low abilities (高分低能)" students (Wang, 2007, p. 93). These students were rote learners, lacked independent thinking ability, spent too much time memorising for exams, and could not learn by themselves. Problems in English language teaching were severe, as knowledge-based and teacher-centred classroom teaching could result in students' lack of ability to use the language for communication, which failed the nation's aim of cultivating competent language users (Wang, 2007). Wang (1999, p. 47) summarised several problems that emerged after the implementation of the 1988/1996 syllabi, which are listed as follows:

- The development of language competency was overlooked;
- There were overlaps between the curriculum of primary schools and secondary schools, which caused a waste of resources and damaged students' motivation;
- The vocabulary requirement was not adequate for students to develop the four language skills: listening, reading, writing and speaking;
- The evaluation of learning was heavily based on paper-and-pen tests;
- Teachers varied considerably in their language proficiency and teaching ability.

To address these problems, the Ministry of Education issued the 2001/2003 curricula and made further changes in the 2011/2017 versions. As this study will focus on English teaching and assessment in the middle school context, I will focus on introducing and discussing the 2001/2011 versions. The 2011 version can be considered the final version of the 2001 trial document, with a sole focus on compulsory education (Gu, 2012). It further emphasised that English should be taught as a communicative instrument and as a medium that connects different cultures (Gu, 2013). The content and requirements remained almost the same as the 2001 version, despite the reduced difficulty level in language skills, given the actual level Chinese students can reach (Gu, 2013; Zhang, 2012).

The new curricula adopted the principles of quality education and aimed to develop both students' language knowledge and practical language abilities (Wang, 2007). There were four fundamental changes in the new curricula regarding the objectives, content, methods, and assessment. Firstly, the overall objectives of the curricula moved from teaching language to educating students through the experience of language learning. Students' overall ability in language use would be promoted through the comprehensive development of five elements: language skills, language knowledge, learning strategies, cultural understanding, and learning attitude. Language skills involve listening, speaking, reading, and writing skills. Together with language knowledge, which comprises phonetics, grammar, vocabulary, and linguistic functions, they build the foundation of overall ability in language use. Learning strategies include communicative, resourcing, cognitive and metacognitive strategies, which help to improve self-learning ability and learning efficiency. Cultural understanding involves cultural knowledge, understanding, and awareness, which can facilitate the appropriate use of a foreign language. Learning attitude, which incorporates international perspectives, confidence, and motivation, promotes learner's autonomy and continuous development. The new changes have placed language competency in a vital position, with communication and language knowledge sharing equal importance. In addition, language was no longer the only focus of the curriculum – the cultural knowledge behind and the learning strategies within language learning have also been brought up to match the communicative language teaching approach proposed by the Ministry of Education.

Secondly, the curriculum unified primary and middle school English into one continuous entity to avoid the problem of resource wasting and divided English language teaching and learning into five competence-based levels (see Table 1-1): levels 1-2 correspond with

primary grades 3-6; levels 3-5 with middle school grades 7-9. This is not a mandated plan for every school because students in some regions start learning English from Grade 1, and some in remote regions might start learning English from Grade 7. Schools and students are given the flexibility to follow the system progressively regardless of when they started learning.

Primary school	Grade 3-4	level 1
	Grade 5-6	level 2
Middle School	Grade 7	level 3
	Grade 8	level 4
	Grade 9	level 5

Table 1-1 The design of five competence-based levels in the 2011 English Curriculum

Thirdly, performance descriptions were given for each competence-based level. The descriptions used 'can do' statements fashioned after the Common European Framework of Reference for languages (CEFR) to shape concrete goals in teachers' and students' minds. For example, at level 3, the following descriptions were listed for writing skills (Ministry of Education, 2011):

- Can use commonly used punctuation marks accurately;
- Can design simple charts and posters to deliver messages;
- Can write or reply to simple greetings and invitations according to examples;
- Can use phrases or sentences to describe a series of pictures and write a simple story.

Fourthly, a new assessment system was promoted, attempting to shift the exam-based system to a more progress- and performance-based one. Recognising that the traditional exam-based assessment system played a negative role in students' overall development, the Ministry of Education proposed a diversified system, which combined both formative and summative assessment (Guan & Meng, 2007; Wang, 2007). At the primary school level, the use of formative assessment was encouraged during students' learning progress according to the performance descriptions given, while at the middle school level, both formative and summative assessment were recommended to evaluate students' achievement. The 2001 curriculum listed eight guidelines for assessment (Ministry of Education, 2001):

- Assessment should be student-centred;
- A diverse and flexible assessment system should be built;
- Attention should be paid to the role of formative assessment in student development;
- Summative assessment should focus on assessing students' overall language skills;
- Attention should be paid to the feedback effect of assessment results on teaching;
- Assessment should be used to boost the motivation and participation of students in primary schools;
- Attention should be paid to the relationship between teaching and assessment;
- Assessment for each level should be designed according to the curriculum objectives.

These guidelines indicated that the 2001 curriculum considered that students have a central role in assessment, and a more balanced assessment system should be established. The curriculum designers' views were consistent with Black and William's (1998a) view on the importance of students' role in assessment, and recognise the influence assessment has on students' attitudes, motivation, and self-esteem (Broadfoot et al., 1999). Such a change of view in assessment has a significant impact not only on the curriculum itself but also on the updated curriculum that comes later.

The 2011 curriculum made some changes in the assessment guidelines by rephrasing. However, the principles remain as the two curricula emphasise using formative assessment at the primary school level and striking a balance between formative and summative assessment for middle school students. The assessment guidelines in the 2011 curriculum are similar to those in the 2001 curriculum (Ministry of Education, 2011):

- Assessment should be used to make positive guidance to stakeholders;
- Assessment design should reflect the centrality of the students;
- The content and criteria of assessment should be based on the curriculum objectives;
- A diverse and rational assessment system should be built;
- Formative assessment should be used to monitor and promote the process of teaching and learning;
- Summative assessment should focus on assessing students' overall language skills;
- Attention should be paid to the relationship between teaching and assessment;
- For the primary school level, assessment should be used to motivate students;
- Middle school graduation tests should be appropriately designed and implemented.

Despite the similar guidelines, the 2011 curriculum standards provide teachers with extra practical guidance on pedagogy and assessment. For primary schools, the 2011 curriculum suggests that teachers should use formative assessment as the primary assessment method and adopt summative assessment as a supplement. Teachers should also provide instant feedback to students in class and design assessment tasks similar to everyday teaching activities (Ministry of Education, 2011). Nineteen exemplars are provided for primary school practice. All the exemplars demonstrate the application of scaffolding, such as pictures, graphs, hint words, and pre-given sentence structures, to support students through the learning/testing process. No summative assessment example is provided, suggesting the curriculum designers' effort to promote the use of formative assessment at this level.

For middle schools, exemplars for formative and summative assessment are given separately. Examples of summative assessment are test forms commonly seen in informal and formal exams: multiple-choice, cloze, reading comprehension, true or false, article writing, and speech making. On the other hand, the examples for formative assessment are more associated with practice in classroom settings. The following techniques are mentioned in the documents: classroom observation, classroom questioning, discussion, portfolio, self- and peer assessment, small quiz, and one-on-one tutorial. The curriculum standards stress that teachers should apply various pedagogical techniques to achieve information about students' learning progress. They should record the messages in time and provide instant, concrete feedback to assist students in closing the gap between their achievement and learning goals (Ministry of Education, 2011). The examples reflect the core value of 'quality education' and demonstrate the combination of both formative and summative assessment. However, detailed explanations about how teachers should use the techniques to improve students' learning are not provided. They only inform teachers about the tools but not how to use them.

When the curriculum policy puts the overall development of children as a priority and put selection in a minor position, an assessment system that uses formative means to improve teaching and learning should be in place. However, the reality is that the high-stake examinations, such as Gaokao and Zhongkao, are, to many people, the fairest way of gaining an opportunity for further education and, to some extent, success. Teachers tend to teach to the test, and students have less interest in what is not tested, especially when they come close to those high-stake examinations that could decide their fate. Therefore, exam-oriented

expectations have become a stumbling block that keeps ‘quality education’ from being successfully implemented, and curriculum designers avoid this tension in the documents, as this is something beyond their control (Gu, 2012).

All in all, changes made in the 2011 curriculum are major steps forward from the 2001 versions. The 2011 curriculum has given rich examples of assessment activities for teacher’s reference, providing some exemplars for teachers to follow. The only example given in the 2001 curriculum was a self-evaluation form, but it did not seem like a formative assessment practice because feedback was not involved in the instructions. Despite the progress made, the strain between the vision of a balanced assessment system and the reality remains, and leaving the tension between a high-stake assessment system and ‘quality education’ principles unchallenged is not a good recipe for this confusing situation. More changes need to be made regarding the current assessment system and the perception and application of teachers, students, and other stakeholders about formative assessment.

1.4 Research problems

According to Liu and Xu (2017), teachers’ assessment activities can be influenced by several factors, including prior experience in their education, power relationships in their workplace, and the contexts where assessment takes place. These factors also affect the implementation of curriculum and assessment policy, as teachers are the essential personnel in delivering the vision of assessment reform into classrooms (Davison, 2004). Although there have been curriculum and assessment reforms over the last two decades in China, there remains a mismatch between what the government sees as an educational vision, what permeates the curriculum and assessment system, and what is happening in classrooms, especially in terms of formative assessment (Gu, 2014). The cause for the mismatch cannot be boiled down to the misconduct of teachers, and the problem should be interpreted from three perspectives: teachers’ assessment activities, their understandings regarding assessment, in particular, formative assessment, and how they understand and enact assessment policies in their context, of which the reasons are elaborated below.

As a relatively new concept in China, formative assessment has only been introduced to the Chinese education system in the last two decades (Liu & Xu, 2017). Most teachers and school leaders are accustomed to the traditional exam-oriented assessment system and have

no personal schooling experience in the new assessment model. These educational stakeholders need to adjust their classroom practices and roles and respond to the new assessment reform through their worldviews. From the background of the study, it is evident that two competing assessment discourses coexist in the English language classrooms in China: one lies in the long-held examination-oriented history, and the other lies in the innovations that introduce progressive and interpretive practices of assessment (Liu & Xu, 2017). This inevitably involves teachers as the agents who implement formative assessment and affects their perceptions of this new concept. As a result, teachers' understandings regarding assessment should be studied. The experiences that shape their thinking should also be explored and analysed.

Teachers' assessment activities should also be investigated. Empirical studies of teachers' formative assessment activities in China are few and far between, especially in the basic education setting. A few studies on formative assessment in basic education have identified some popular methods teachers used. As Zhang (2013) reported in a research programme involving 72 high schools from 8 provinces, there were seven types of commonly performed formative assessment activities: classroom observation, peer- and self-assessment, praise and encouragement, group study, classroom competition, portfolio, and classroom interaction. However, at the same time, many teachers are reported to have difficulties understanding what formative assessment means. Zhao (2013) lists three types of common misunderstandings regarding formative assessment from teachers' perspectives: equating formative assessment with praise and encouragement, which overlooks the need for effective and instant feedback; taking in-class quizzes as formative assessment, which focuses on summative results; and the negative feeling that formative assessment is not practical. The reasons why teachers choose and how they understand different assessment methods could be complex and situated (Liu & Xu, 2017). It is not easy to judge whether their activities are qualified, but it is possible to understand their choices within the context of their practices. Therefore, the relationship between teachers' assessment activities, working environment, and assessment values should be explored.

Assessment policy is also an essential part of the context that needs to be investigated. As a concept brought up and developed in the Anglophone context, formative assessment has been borrowed from China's "reference society" to review and reform its educational system (Tan, 2016, p. 195). In theory, its function of improving learning and education quality suits

the nation's strategic development of producing all-around learners (Chen, 2017). However, a loose coupling has emerged between the policy and practices, and the endeavour to establish an effective enactment of formative assessment was found to be complex and problematic (Chen, 2017; Steiner-Khamsi, 2014). Tan (2016, p. 196) introduces the notion of "theory-ladenness", which refers to the situation that what people perceive and observe is under the influence of one's prior experiences, beliefs, and values, to explain why this situation would appear when formative assessment, a borrowed idea, is translated into a different context. Her idea corresponds with Chen's (2017) report, which indicates that in many cases, formative assessment in Chinese classrooms is "decontextualised, recontextualised and culturally appropriated" (p. 75). When the new assessment policy comes into place, teachers are likely to incorporate new assessment methods in their classrooms with some adaptation to meet the local needs. How teachers interpret current assessment policy should be explored. How teachers design new assessment methods according to the new assessment policy and the local demands should also be given more thorough attention.

1.5 Motivation for this study

My motivation for this study emerged from my personal experience as an English language learner, a language teacher, and a curriculum designer in an online educational company. I started learning English in grade one in primary school. Throughout the years of my basic education, English has been one of my main subjects. After attending the Gaokao in 2011, I went to university and chose English as a major for my Bachelor's degree. Looking back at my English learning journey, I find that assessment has always been an important part of my study. Mid-term tests, final exams, and quizzes have never been absent, and the memories of the high-stake exams I took, such as Zhongkao and Gaokao, are still fresh. Assessment has been a source of my motivation for English learning and also a source of pressure, and I notice the same feelings when I teach my students. For years I have heard about the debates about whether test-driven education is beneficial to students, and the conversation usually ended with a statement that 'if the assessment system remains the same, there is little we can do'. Before coming across the concept of formative assessment, my understanding of assessment was limited to the impression that assessment equals testing. It was after my Master's study and my teacher qualification training that I realised there are other forms and purposes of assessment.

I realise the potential and benefits of formative assessment during my work experience as a language teacher in different institutions and a curriculum designer in an online educational company. A part of my job responsibilities was planning classroom and formal assessment and designing online English courses and materials for other teachers. During the collaboration with my colleagues, I found that two of the limitations of language education, both online and offline, were the lack of continuous monitoring of students' progress and continuous support for their learning. Although exercises and exams were usually developed for students to check on what they have learnt in class, they were not efficient enough to provide students with sustainable feedback and support throughout their learning. To tackle these issues, I tried designing questions in the middle of the lesson so that students' progress could be regularly checked on. I cannot say that this solved the problems immediately, but I received some positive feedback from the teachers and students, which informed me that such an approach could be beneficial.

I took a step further when I planned this study. In our schooling system, tests and quizzes have been used frequently to boost students' learning. However, these assessment forms usually occur at the end of a learning process. What happens during the learning process is unknown, and overusing testing may result in the loss of opportunities to assess and help students efficiently throughout the learning procedures. With this study, I hope to have an opportunity to observe the current assessment activities performed in today's English language classrooms and talk to the teachers about their understanding of assessment. I hope my study can create a more fruitful discussion on how we can use assessment more efficiently in teaching English in Chinese classrooms.

1.6 Aim of this study

I intend to explore teachers' assessment activities, their understandings of assessment, particularly formative assessment, and their enactment of the current assessment policies in middle school English classrooms in Shenzhen city, China. The first reason for choosing this context was that, while Gaokao received massive attention from educational researchers in China and overseas, Zhongkao, which middle school graduates take for their promotion to high schools, witnessed more pressure experienced by students due to the limited and unevenly distributed public high school educational resources and received less academic spotlight. The second reason was that the city of Shenzhen was experiencing a Zhongkao

reform when this study was conducted, which made the study timely and important. The following research questions are proposed for studying the issues at hand:

RQ1: What classroom assessment activities do teachers adopt? Why do they choose these activities?

RQ2: What are teachers' understandings of assessment, in particular formative assessment? Why do they adopt such beliefs?

RQ3: How is the current assessment policy reflected in the textbook, teacher guidebook and the Zhongkao teachers work with? What actions have the teachers taken to put the assessment policy into practice under the influence of these materials? Why do they enact the assessment policy as such?

Chapter 2 Literature Review

To understand where the research focus of this study is situated within broader scholarship, this chapter reviews the existing theoretical and empirical literature relevant to this topic. Firstly, the debates around and practices of language theories will be discussed (section 2.1). The development of language theories has influenced the development of thinking regarding curriculum, pedagogy, and assessment in language classrooms. To realise the theoretical foundation of today's debates in assessing language learning, it is helpful to retrace the history of development in language theories. Secondly, the development of language assessment practices will be reviewed (section 2.2). Language assessment is closely related to its informing language theory. Their relationship will be carefully explained, and the role of assessment in language classrooms will be given special attention. Thirdly, studies of teacher enactment of language assessment policies will also be discussed (section 2.3). Theories and policies might provide a direction for teachers' practices, but policy enactment will finally decide the classroom practices. Studies into the factors influencing policy enactment of language teachers will, therefore, be reviewed. More particularly, studies related to the enactment of language assessment policies will be examined.

2.1 Language theories: debates and practices

During the past century, language teaching and learning theories have reflected developments in linguistics, psychology, sociology, and pedagogy (Mitchell & Myles, 2004). The ongoing debates have brought changes in dominant teaching and learning theories and in the classroom practice that follows. The literature pertaining to the shifting sands of language pedagogy will be presented in three parts: behaviourist approach (section 2.1.1), in which language learning was considered a formation of habits that could be obtained through mimicry and memorisation (Arnfast et al., 2010; Lightbown & Spada, 2013); communicative approach (section 2.1.2), which stemmed from the critiques of behaviourism and embrace the importance of communication and context in language learning (Mitchell & Myles, 2004); and sociocultural approach (section 2.1.3), which turned researchers' focus from viewing language learning as a development within an individual to viewing it as a socially mediated process (Arnfast et al., 2010; Lightbown & Spada, 2013).

2.1.1 Behaviourist approach

Gaining popularity in the first half of the 20th century, behaviourism remains a prominent learning theory, and practices informed by behaviourist thinking widely exist in today's language classrooms (Lightbown & Spada, 2013). Behaviourism stemmed from the thinking in psychology that viewed learning of any kind as being based on stimulus and response. From a behaviourist point of view, in an environment where a number of stimuli are provided, if humans respond successfully and achieve the desired outcome, their response would be reinforced and later become a habit (Mitchell & Myles, 2004). Language development was considered a formation of habit – a certain response was needed in a certain situation, just as a greeting was needed when you met someone. It was also argued that when people learnt a foreign language, they would start with the habits they formed in their first language, and that these habits could interfere with their learning of a foreign language (Lightbown & Spada, 2013).

Under the framing of such a theory, effective teaching was usually associated with the motto of 'practice makes perfect' and the mastery of differences between languages. According to Brooks (1966), one of the proponents of behaviourist theory in language learning, foreign language should be learnt through the process of "model, reward and reinforcement" (p. 358). Students learnt by adjusting their language behaviour in accordance with the appropriate models and repeating the process often enough until the sound and word choices were correct (Brooks, 1966). Another proponent of this theory, Lado (1957), believed that the differences between the native and foreign language systems were the source of trouble in foreign language learning. To successfully teach a foreign language, teachers need to compare the patterns of native and foreign languages to predict the learning problems (Lado, 1957). Work in this tradition has provided theoretical grounds for many foreign language teaching methods, such as the audiolingual method, which attempts to use the stimulus-response-reinforcement model to create habits in language learners, and the PPP method (presentation, practice, and production), which is a variation of the audiolingual method that put learners in situational contexts (Harmer, 2007b).

Despite its continuing influence, behaviourism encountered many criticisms. From the 1950s, linguists shifted their focus from the description of structure and corpus of languages to the generative nature of languages (Mitchell & Myles, 2004). An example of the clash of ideas

about how languages are learnt was Chomsky's (1959) review of Skinner's (1957) book 'Verbal Behavior'. While Skinner explained in detail his behaviourist view on learning languages, Chomsky fiercely critiqued this view by arguing that children did not acquire their verbal and non-verbal behaviours only through "meticulous care" from adults (1959, p. 42). From his perspective, all human beings inherited innate knowledge about language in their minds, which drove their language learning (Chomsky, 1981). Such innatist and mentalist views of language share a distinct feature from other language theories reviewed in this study, such as the sociocultural learning theory informed by Vygotsky's work (section 2.1.3)

Despite the criticism, behaviourist theory continues to be widely practised. Even to this day, many language courses adopt methods such as the audiolingual method, using drills to acquaint students with pronunciation and grammar of a second language (Brown & Lee, 2015). Behaviourist theory has offered an angle of looking at language studies, but it should be noted that no theory could be taken as the whole story for something as complex as language. Other ideas from both the field of linguistics and the field of language teaching addressed the complexity of language and language learning and gave rise to discussions regarding teaching and assessing language for communicative purposes. These ideas, which contributed to the theory and application of the 'communicative approach', will be discussed in the following section.

2.1.2 Communicative approach

Following the criticism of behaviourism, the importance of communication and context in language learning received wider recognition with the development in both field of linguistics and field of language teaching. Informed by anthropologists such as Hymes (1972), whose work explored how people used languages to communicate in acceptable ways in different contexts, language educators and applied linguists gave more attention to language use in context (Leung, 2022). Other researchers explored the functions of language in situational contexts and how meanings were expressed through grammar, in ways that have contributed directly or indirectly to communicative language teaching. For instance, Halliday and Matthiessen (2000) introduced the concepts of field, tenor, and mode to describe the functional relationship between human experience and language expression. They defined field as the "culturally recognised repertoire of social practices and concerns",

tenor as the “culturally recognised repertoire of role relationships and interactive patterns”, and mode as the way “linguistic resources are deployed” in the given context (Halliday & Matthiessen, 2000, pp. 320-321). Together, these contextual variables define the “‘ecological matrix’ in which particular types of text are processed”: the field relates to ‘what activity is going on’; the tenor refers to the people who are taking part; the mode concerns the part language is playing in the context. They make it possible for speakers to choose the right register, which refers to the specific lexical and grammatical choices made, to appropriately convey their meanings according to the events they are involved in, the people they interact with, and the kind of linguistic resources that can be deployed (Halliday & Matthiessen, 2000, p. 321).

Although Hymes, Halliday, and Matthiessen’s focus was not on language learning but on language as social behaviour, many educational practitioners in the field of language teaching took their views as their theoretical basis, which gave rise to widespread practices of the communicative approach (Savignon, 1991). The communicative approach emphasises a comprehensive view of language competence, which gives special attention to actual language use (Littlewood, 2011). It encourages authentic and functional use of language for meaningful communication in a real-life context (Brown & Lee, 2015). From such premises, language proficiency has been redefined to represent language learners’ competence at not only grammatical level but also sociolinguistic, discourse, and strategic levels (Leung, 2022), and language learning has no longer been viewed as mechanical habit formation. Instead, it has been understood as resulting from the interaction between language users, involving collaborative creation and negotiation of meaning, learning through feedback received, and the process of trying out different ways of saying things (Richards, 2006). The social dimension of language use, in particular, the appropriateness of utterances, becomes the focus of language teaching (Leung, 2022). Classroom activities such as group work, role play, and project work are favoured over memorising dialogues and drills. Learners are asked to participate in the classroom on a more active and cooperative basis, and teachers are asked to take the role of a facilitator and monitor of the class rather than being the model for correct speech and writing (Richards, 2006).

Empirical studies have explored how the communicative approach works in classroom settings. In her research project, Savignon (1972) took an approach using communicative strategies in a beginner French programme, which was different from the dominant approach

of pattern practice and error correction at that time. Her study included three groups with a total number of 42 students. One control group was assigned to use the audiolingual method in French teaching, while two experimental groups used the communicative teaching approach. A written test and a speaking test were administered at the end of the programme. According to the test results, the students taught through the communicative approach achieved a similar level of accuracy in the written test and a higher level of fluency and comprehensibility in the speaking test, compared to those taught with traditional pattern drills. Therefore, Savignon (1972) argued that communicative teaching could improve communicative competence. While Savignon (1972) advocated that even beginners could respond well to communicative teaching, Canale and Swain (1980) made different claims by arguing that, without mastering a certain level of grammatical knowledge, second language learners, especially those at their beginning stage, will most likely demonstrate a restricted ability to express in language and be unable to pay attention to using language in a task, despite that their lack of grammatical knowledge could be compensated by their strategic competence such as gestures (Canale & Swain, 1980). However, their view contradicted Hymes' and Halliday's opinions, as Hymes (1972) argued that grammaticality was not the primary concern of effective communication, while Halliday (1985, p. xvii) believed that meaning was not naturally related to grammar, as the evolution of human language began with no grammar at all.

There have also been other criticisms of the communicative approach. Firstly, Howatt (1987) argued that the communicative approach adopted major principles and features from earlier methods such as direct method and audiolingual method. Its uniqueness was based on its communicative activities rather than its theoretical underpinnings (Kumaravadivelu, 2006). Secondly, Kumaravadivelu (2006) mentioned the difficulty in delivering authentic communicative language teaching in many classrooms. Although the communicative approach claimed to provide authentic communication that characterised interaction in real-life settings, Kumaravadivelu argued that it was doubtful that a communicative curriculum could guarantee meaningful communication in classrooms. Even if teachers claimed to have followed the communicative approach, they could still fail to create opportunities for genuine communication. Thirdly, the communicative approach could have trouble adapting to different cultures and contexts of language learning worldwide (Harmer, 2007b; Kumaravadivelu, 2006). Factors such as classroom size, educational cultures, and teachers'

beliefs about communicative language teaching could all affect the practice of communicative approach in real classroom settings.

Sato and Kleinsasser (1999) conducted a study that reported on second language teachers' beliefs, knowledge, and practices regarding the difficulties in implementing the communicative approach. Ten high school Japanese teachers in Australia participated in the study. Open-ended interviews, classroom observations, and surveys with teachers were carried out for data collection and analysis. The study highlighted how the teachers dealt with the communicative approach in language classrooms. The teachers wanted to teach Japanese for communication, with speaking and listening given a high priority. However, they showed reluctance to implement interactive or innovative activities, as they found them time-consuming to plan and perform. They also pointed out that communicative language teaching involved little grammar learning. While guidelines, scholars, and policymakers suggested that grammar was not the core element in communicative teaching, most teachers believed grammar was essential for language learning (Sato & Kleinsasser, 1999). Sato and Kleinsasser (1999) used multiple data resources to present teachers' perspectives on communicative language teaching. Their findings echoed those of Kumaravadivelu (2006) on the difficulties in understanding and implementing the communicative approach and also raised questions on how teachers could transform their language classroom into a more communication-oriented one.

Sato and Kleinsasser (1999) also reported interesting findings on teachers' views on assessment in communicative language teaching. According to the teacher participants, the assessment tasks offered under the communicative teaching curriculum resulted in slight obstacles in their teaching. Firstly, while speaking and listening were emphasised in communicative teaching, government guidelines for communicative assessment involved all four skills, which seemed to be given equal weighting. This suggests that although communicative language teaching was promoted in this setting, there has been an alignment issue regarding the pedagogy and assessment – the language was taught in a way that highlighted listening and speaking skills, but it was not assessed in the same way. Secondly, some teachers were reported to be in a dilemma, as according to the guidelines, they were not allowed to conduct grammar tests. They thought that grammar tests could be an effective way of improving students' writing. However, these tests were considered by the curriculum guidelines not communicative enough. All they could do was provide authentic materials for

students to read, which was a valuable practice but took more time. Sato and Kleinsasser (1999) noted that, in this study, teachers would allow their understanding of the policy to outweigh their practice of communicative language teaching. The authors also argued that if the policy does not provide better grounds for alignment between pedagogy and assessment, it is doubtful that a communicative curriculum could guarantee meaningful communicative teaching in classrooms.

Deng and Carless (2010) conducted another study focusing on the influence of examinations on implementing the communicative approach in the Chinese context. The case study involved four English teachers from two primary schools in Guangdong province, where local schools were encouraged to try out the communicative approach promoted in the new English curriculum, which the Ministry of Education introduced in 2001. The study put an emphasis on the role of examination. Through lesson observations and interviews with teachers and school principals, the authors intended to find out how teachers interpret the relationship between pedagogy and examination and whether examinations have been a constraint to implementing the communicative approach.

According to Deng and Carless (2010), the principals and teachers in the two primary schools showed different attitudes and understandings regarding the innovative approaches promoted by the new curriculum, which led to different interpretations and practices. The principal and teachers in School A favoured traditional teaching and allocated more time for exam preparation. This was due to the non-communicative nature of their internal and external exams and the accountability pressures of school ranking. The teachers have also shown a lack of knowledge and confidence in communicative language teaching, which contributed to their negative attitudes towards new practices. On the other hand, School B was optimistic about the new approach. The principal of the school advocated new approaches derived from the communicative approach, such as story-telling teaching and task-based approach, and the teachers were also observed to actively integrate communicative teaching in their classrooms. However, it should be mentioned that school B was not free from the influence of examinations. One teacher in school B reported that the external public exams have become more difficult than usual, and he felt the need to reduce communicative teaching and increase test drilling. Examinations have been observed as an important role that influenced teachers' values and beliefs in both cases, and teachers' beliefs

and pedagogy have further influenced the interplay between communicative teaching and exam-oriented tradition.

In the study, Deng and Carless (2010) argued that although the role of examinations was an impeding factor in implementing communicative teaching, teachers' beliefs could be an even more powerful obstacle to innovative pedagogy. I would instead question whether this is only related to teachers themselves as individuals. Teachers' practices are closely related to the educational system and the context they work in. The examination-oriented setting has an impact on how schools and students are assessed, which is closely related to test performance; the accountability pressure of schools has an impact on school policy regarding how examinations are viewed and used; principals and teachers are encouraged to carry out approaches promoted by the new curriculum according to their own interpretation, in which examination is a factor that they could not avoid; and finally, teachers end up with having to put more emphasis on exams, willingly or unwillingly, and with the exams not focusing on communication, communicative approach has to be put aside. Deng and Carless (2010) also argued that examinations should be shifted to a more communicative direction to give teachers the incentive to teach communicatively. As the internal and external exams for the two schools were non-communicative, building an alignment between communicative teaching and assessment could be a helpful direction. Nevertheless, as the authors have also mentioned, innovation and examination have a complex relationship. If teachers' beliefs and the educational system are not changed, it is difficult to make fundamental changes happen.

All in all, the communicative approach views language as social behaviour and language learning as a result of meaningful interaction and collaboration between language users. It encourages authentic communication in a real-life context and has received support from many educational practitioners. However, it has also received criticism for application difficulties and encountered alignment issues regarding pedagogy and assessment in certain contexts. Another theory that views language learning in social terms is the sociocultural theory. Theorists in this field claim that language interaction should not be taken simply as a source of input but should be viewed as playing a more central role in learning. They have also offered new perspectives on language assessment. The following section will review this strand of thinking and research in second language learning.

2.1.3 Sociocultural approach

Since the 1980s, second language learning research has been associated more with sociocultural theory, which builds from the work of Lev Vygotsky, a Russian psychologist (Lantolf & Thorne, 2006; Mitchell & Myles, 2004). Vygotsky (1978) argued that the most significant factors that accelerate human's cognitive development are the interactions within different social environments, such as interactions with family, peers, teachers, and workmates. As opposed to the innatist view of language acquisition, which believed language faculty is biologically endowed, or the functional perspectives on language learning, which viewed thinking and speaking as two related but independent processes, Vygotsky's theory considered that language developed primarily from social interaction and that thinking and speaking are deeply interrelated (Lightbown & Spada, 2013). Language is a tool for thought and a means of mediation (Vygotsky, 1978). Through language, people can direct their attention to important information and problem-solving; in turn, language as a cognitive tool can also shape their ways of thinking (Mitchell & Myles, 2004).

Vygotsky's work mainly focused on children's psychological development, but second language theorists have taken up several key ideas from the interpretation of his work. One of the key concepts that are frequently mentioned is the Zone of Proximal Development (ZPD). Vygotsky defined the concept as:

> It is the distance between the 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. (Vygotsky, 1978, p. 86)

Another well-known concept is the metaphor of scaffolding. Scaffolding refers to the process of supportive conversation in shared activities that promotes learner's internalised knowledge (Mitchell & Myles, 2004). It relates to the assistance provided by an expert to a novice, which helps the novice carry out tasks and activities with guidance, and finally leads to development within his or her ZPD.

Besides these two concepts, Vygotsky's argument of what counts as good assessment has also been taken to understand assessment from a sociocultural perspective. In building the concept of ZPD, Vygotsky (1978) criticised the educational practices that assessed development and guided educational assistance on the sole basis of one-off assessments of

individual performance. He argued that two developmental levels of a learner should be taken into account: the actual developmental level, that is, the level a learner can reach by solving the problem himself; and the potential developmental level, that is, the level the learner can reach with assistance from adults or more competent peers. According to Vygotsky, the potential developmental level could differ from a learner's actual development. The potential developmental level is more indicative of the progress made by the learner, while actual performance only suggests what the learner can do when assessed. This suggests that sociocultural theory is promoting a new way of thinking about assessment that focuses more on the developmental process of learners, which is different from the view of how assessment has conventionally been considered.

Following Vygotsky's thinking, a group of sociocultural studies has been conducted to examine second language development and assessment. Aljaafreh and Lantolf (1994) conducted a pioneering longitudinal study on three adults who learnt English as a second language. The participants were involved in weekly one-on-one tutorials for eight weeks and received feedback from their tutors on their writing assignments. At each tutorial, they were first asked to re-read their writing to look for errors without assistance. Then, the tutors would work with them to help them find their mistakes. If a mistake were identified, the tutors would scaffold the participants to correct the mistake. The study focused on four grammar points that were considered to be within the participants' ZPD. The researchers looked for evidence of an increase in accuracy in using the four grammar points over time and evidence of participants' developing capacity for self-correction if mistakes continued to appear in their writing.

In the data analysis, Aljaafreh and Lantolf (1994) focused more on the process of correction and negotiation between the participants and tutors instead of on how well the students performed at the end of the study, which reflects the sociocultural perspective of development and assessment. They explained their findings from three perspectives. Firstly, they reported that collaboration in the correction process could enable learners to understand a feature of the second language (L2) that had not been understood successfully. They also argued that the simple act of building a collaborative relationship could be an effective way for tutors to promote and monitor their students' studies. Secondly, they suggested that different learners have different ZPDs for the same target knowledge. The learners may make different mistakes regarding the same grammar points, but making the same mistake does

not necessarily indicate that the learners have the same problems in learning. The authors further argued that it is vital for teachers to assess both the learners' actual and potential developmental levels and to give different learners different feedback. However, they only provided assessment and feedback examples that took place in one-on-one tutorials and did not address how customised assessment and feedback could be provided in classroom teaching.

While Aljaafreh and Lantolf (1994) focused their study on the collaboration between a novice and an expert, another strand of sociocultural research on second language learning focused on the interaction between peers in language classrooms. Ohta (2000) conducted a study on two adult learners in a university-level Japanese language class. She aimed to investigate how the two learners interacted with each other in order to mutually provide developmentally appropriate assistance. The learners were asked to collaborate in an oral translation task with high demands for form and grammatical accuracy. The first learner, Hal, had higher proficiency in Japanese, while the second learner, Becky, had greater difficulty in completing the task. When Becky struggled with the task and indicated that she needed assistance with interactional cues, Hal would provide timely help and intervene with the linguistic errors Becky made. Ohta (2000) claimed to have observed in the study that Becky improved dramatically in her use of unfamiliar grammar points and sentence structures. In addition, she reported that Hal had also developed his language proficiency through the process. As a result, she argued that the data involved evidence of learners providing scaffolding for one another to produce language utterances that they could not produce themselves.

Similar to Aljaafreh and Lantolf's (1994) study, Ohta (2000) concentrated on the importance and production of collaboration; only in this case, her study shed light on the support provided by more capable peers. Both of their research approaches were inevitably restricted by "some of the usual difficulties in developing causal explanations and generalisations through naturalistic research" (Mitchell & Myles, 2004, p. 219). As Mitchell and Myles (2004) commented, there might be other factors that contributed to their improvement, for example, the continuing exposure to the language and the amount of time and effort they spent in learning a language. However, despite the limits in generalisation, both studies offer an interesting reflection on sociocultural aspects of assessment. Unlike the studies discussed in the sections regarding other language theories, both studies make no use of written or

spoken tests in their research design for data collection. They both draw on the process of the conversation between the participants to make judgements about their performance, and focused less on the results of formal tests. This could be interpreted as an alignment between sociocultural pedagogy and assessment, as sociocultural approach emphasises development made during social interactions. It does not suggest that sociocultural theory has less concern over how learners perform in formal assessment or how interaction and collaboration could improve learner's performance in formal assessment. Rather, it suggests that assessment can be carried out from different perspectives, and result is no longer the only primary concern.

So far, this chapter has focused on how different theories view language and how these theories influence thinking regarding language curriculum, pedagogy and assessment. Behaviourist approach viewed language learning as a habit formation that could be achieved through memorising linguistic elements and drilling different language skills; communicative approach advocated a more comprehensive view of language competence, in which actual language use in an authentic context was encouraged; sociocultural approach shifted the focus from viewing language learning as development achieved by individuals to viewing it as development achieved through social interaction and cooperation. These theories have influenced not only the development of language pedagogy but also the development of language assessment. The alignment issue between pedagogy and assessment has also been a critical concern in language teaching and learning, as practitioners seek to teach and assess language in unified ways, despite that few theories address the interconnection between learning, teaching, and assessment in language classrooms. In the next part, a history of the development of language assessment will be introduced, with a review of how language theories influenced the development of language assessment and how different assessment practices interplay with curriculum and pedagogy.

2.2 The development of language assessment

Historically, language assessment has been influenced by understandings of the nature of language and the development of language teaching and learning theories (Ingram, 1985). Common language test forms (section 2.2.1) and classroom language assessment (section 2.2.2) have evolved over the years and have played a considerable part in shaping language classroom practices. Before the 1960s, language testing, informed by behaviourist theories, focused more on individual language elements, which made discrete-point test a popular

choice (McNamara, 2013). In the 1970s and 1980s, communicative theories of language teaching informed the emergence of the integrative test, which added sociolinguistic aspects to testing theories, such as the knowledge of how language is used in a social context (Spolsky, 2008). In the 21st century, the connection between learning and assessment has been given more attention, and sociocultural theory of learning has provided a foundation for constructing different forms of classroom language assessment (Katz & Gottlieb, 2012). The development of, and the debate around, each type of language assessment will be discussed below. How each type of assessment has left its trace in and impacted classroom practices will also be addressed.

2.2.1 Common language test forms

In the early 20th century, standardised tests gained their reputation in successfully selecting soldiers for the US army during the First World War (Black, 2001). Celebrated by testing companies and other interested parties, 'objective' tests, influenced by the 'objective' techniques of standardised tests, were popularised in schools as an ideal instrument for educational assessment (Spolsky, 2008). During this period, behaviourism and behaviourist language theory were in vogue. Behaviourism regards learning as acquiring or reproducing the knowledge teachers intend and assessing learning as measuring the mastery of what is taught (James, 2008). At the same time, language was viewed as a multitude of discrete linguistic components which could be learnt via stimulus and response. The two strands of educational beliefs mingled, which led to the production of discrete-point tests, one of the most common types of tests developed in this early period of modern language assessment (Brown & Abeywickrama, 2010). This kind of test was based on the assumption that language could be broken down into individual points, and tests could be conducted on each point respectively (Brown & Abeywickrama, 2010). It was also considered to be able to provide a stimulus to learners so that they could give an appropriate response (Ingram, 1985). Multiple-choice, true-or-false, and spelling tests were popular means of assessment in this manner (McNamara, 2013).

There were two reasons why discrete-point tests prevailed, and both reasons were strongly challenged in later days. The first reason was that reliability and validity were considered highly important in test development. Test forms such as multiple-choice and true-or-false, the common forms of discrete-point tests, were viewed as the only ways to ensure the test

was well calibrated for difficulty and measured what was intended to be measured. Other forms of testing, such as writing and speaking tests, were considered practically difficult and unreliable (McNamara, 2013; Spolsky, 2008). However, the testimony supporting the validity of discrete-point tests was invalidated. As Oller (1979) argued, language competence requires the integration of different skills, and thus additive tests of grammar, vocabulary, and other discrete points could not reflect a person's actual language capacity. A discrete-point test's validity and reliability were therefore dubious.

The other reason that discrete-point tests gained popularity was that the behaviourist teaching practices favour teaching language by components (Davies, 2014). Language was believed to be "a set of systems of structural contrasts at the level of phonology, morphology, syntax and semantics" in the behaviourist view, and the focus of teaching was on the mastery of systematic features of language (McNamara, 2013, pp. 341-342). The teaching method provided the ground for discrete-point tests, which featured separate subtests of individual systems of language. However, such an approach was criticised as requiring a "decontextualization" in language learning and testing (Brown & Abeywickrama, 2010, p. 13). Researchers such as Oller (1979) also argued that language competence should be assessed as a unified set of interactive abilities. As the idea of communicative language learning emerged in the 1970s and 1980s, language was no longer understood to be a bundle of unrelated linguistic elements. Communication, context, and authenticity were increasingly emphasised in language learning, and more attention was given to a new approach in language assessment – integrative testing.

The notion of integrative testing was based on the idea that "the whole of the communicative event was considerably greater than the sum of its linguistic elements" (Brown & Abeywickrama, 2010, pp. 12-13). While the discrete-point test assessed language as habit-forming and a group of discrete linguistic elements, the integrative test viewed language competence as a skill that could be performed in the real world (McNamara, 2013). It corresponded with Hymes's (1972) argument that grammatical competence does not equate with communicative competence and Halliday's (1985) opinion that language is essentially a social behaviour that requires more than linguistic knowledge. It also resonated with the constructivist views of learning, which signifies learning as an active psychological process of making sense of new knowledge and assessing learning as focusing on problem-solving and understanding (James, 2008).

Language tests adopting this approach attempted to reproduce the language use context in a test and had criteria of 'authenticity' for the tasks. This approach has been widely practised in the language testing field, especially in large-scale language tests, such as the Test of English as a Foreign Language (TOEFL) and the International English Language Testing System (IELTS) (McNamara, 2013). Both tests include themes such as university lectures or everyday life scenes because many test-takers are about to spend their life abroad and pursue further study at a university overseas. The tasks resemble the communicative activities they are likely to be involved in and therefore provide evidence for their capability of dealing with those tasks in the future. Language test forms in China have also echoed the call for communicative language teaching and alignment between pedagogy and assessment. An example would be the gradual changes in the English Gaokao. When it was first designed and administered in the 1950s, English Gaokao was a pure discrete-point test covering only discrete linguistic items (Li, 1990). It was re-designed throughout the decades and has become a test assessing both discrete items and integrative use of language (Li, 1990; Qi, 2005). The test designers have been making continuous efforts to promote changes to classroom teaching through new test designs (Li, 1990; Qi, 2005). They wanted to send out the message that the teaching of language use should be given priority in language classrooms, and language skills should be practised in real-life contexts with an emphasis on meaning-making.

Despite receiving considerable support, integrative test was not free from criticism. Some researchers, such as Harding (2014) and McNamara (1996), have criticised integrative tests as too complex and challenging to design, while others have reported that integrative tests were too narrow to explain what counted as competent communicators. Moreover, integrative tests were reported to be having restricted impact on teachers' teaching. Qi (2005) conducted a study in China, intending to find out whether changes in English Gaokao have influenced teaching and learning in classrooms. She collected data through interviews and questionnaires from eight test designers, six English inspectors in the local educational authority, 388 teachers, and 986 students in high schools. Qi's (2005) study suggested that the test designers had fallen short of their goal. Although the testing reform has been carried out over some decades, teaching and learning practices in high schools still focused on isolated linguistic items. Qi (2005) reported that teachers believed what was tested by the English Gaokao was grammar and vocabulary, and the skills needed for the test depended on these basic elements. This belief contradicted sharply with the test designers' intention.

On the other hand, according to Qi (2005), the selection function of Gaokao has pushed teachers and students to work for raising scores. As teachers believed the mastery of grammar and vocabulary would contribute the most to students' test performance, English courses in high schools emphasised discrete linguistic knowledge, such as grammar and vocabulary, instead of the use of language, as promoted by the test designers.

This study illustrates a more complex picture of the alignment issue between pedagogy and assessment. How different groups of stakeholders reflect the ideas behind different test designs has also been addressed. Although test designers have tried to promote new pedagogical and assessment theories through test design changes, teachers' choice of pedagogy and assessment seems little affected. This finding somewhat contradicts with Deng and Carless (2010) argument, as they believed changing the test design might brought changes in styles of teaching. There could be various reasons affecting the enactment of teachers. Firstly, although new theories of pedagogy and assessment were promoted, teachers' values regarding pedagogy and assessment have not changed. This could be related to their prior education or training experience, which gave them few reference models to learn from. Secondly, the selection function of external assessment could convince teachers to hold on to what they are confident in and comfortable with. The pressure from schools, students, and parents would keep them from adopting new theories and methods because of the high stakes of the external test. As a result, the alignment issue of pedagogy and assessment became difficult to solve, and teachers were unable to enact test designers' or policymakers' intentions.

So far, we have discussed how educational and language theories have influenced the development of language testing and reviewed the implication of different language test forms for language teaching and learning. From discrete-point test to integrative test, the focus of language testing has shifted from individual linguistic elements to purposeful and meaningful communication. Different pedagogical theories have affected different language test forms and have also played a part in shaping different language classroom practices. Challenges still exist regarding whether pedagogy and assessment could reach an alignment and whether teacher enactment could be carried out to fulfil the intention for changes. Apart from these tests that often serve as external evaluation tools at specific times in a curriculum, another type of language assessment that should not be neglected is classroom language assessment. As a broader practice of evaluating students while building their language

competence and skills (Brown & Abeywickrama, 2010), the role of classroom language assessment, the focus of this study, will be discussed in the next section, and studies related to classroom language assessment will be reviewed.

2.2.2 Classroom language assessment

In the previous section, I have tracked the development of language testing from discrete-point test to integrative test, the process of which is tightly interwoven with the development of educational and language theories from behaviourism to the constructivist communicative approach. In this section, the influence of sociocultural theory on language assessment will be given more attention. Moving into the 21st century, the practice of assessment is no longer viewed as exclusively related to tests. The relationship between learning and assessment is more tightly drawn, and assessment in language classrooms has been given increasing attention in language education research and educational policies worldwide (Brown & Abeywickrama, 2010; Davison & Leung, 2009). Contemporary classroom assessment draws on sociocultural theory (Katz & Gottlieb, 2012). From a sociocultural perspective, learning requires thinking and actions within a situated context, and assessment should address the process of building knowledge through collaboration (James, 2008). While traditional language tests are usually formally administered procedures aiming at measuring test-taker's performance in a particular domain in a limited time, classroom assessment represents a much broader concept, which can involve teacher's judgments and support of students that are carried out throughout the teaching and learning process (Brown & Abeywickrama, 2010). Assessment is given a more collaborative nature, and teachers are encouraged to collect information from a wide range of assessment tools, which are often integrated into everyday learning activities.

Among the assessment choices available to teachers, a distinction has been drawn between two types of assessment according to their functions and pedagogical purposes: formative and summative assessment. The distinction between the two terms was firstly suggested by Scriven (1967), who argued that summative assessment could be used to assess the outcome of an educational programme, while formative assessment could be used to improve the programme's performance. Bloom (1969) also adopted the same terminology but developed an argument with more respect for students. He suggested that the purpose of summative assessment was to judge what the learners had learnt at the end of a programme, while the

purpose of formative assessment was to provide feedback and support during the teaching and learning process. Over the years, formative assessment has received widespread interest. It has gained growing popularity based on the idea that external testing systems should not dominate classroom learning, and teachers' efforts should be put towards promoting learning instead of promoting test outcomes (Black, 2001). Much research has also been conducted on the practices of formative assessment to find evidence that its development could raise the standard of achievement (Black & Wiliam, 1998a, 1998b).

However, as Black (2001) reported, the development of formative assessment has been repeatedly displaced by an emphasis on summative testing and test results, and its complex relationship with pedagogy has contributed to its slower development. When assessment data, especially test results, are used for purposes such as judgments on school quality and teacher qualification, schools and teachers might take actions to improve students' performance in the measured assessment, which might go against what is initially planned for the students' long-term needs. Although it is suggested that teachers use formative assessment data to build an understanding of students' learning progress and to provide suggestions for students' further development, it cannot be guaranteed that they would do so at the grassroots level. Other researchers also commented that "formative assessment is a central part of pedagogy" (Mansell et al., 2009, p. 9). This could be part of the reason why teachers find formative assessment hard to implement, as it challenges how they teach and think about teaching, learning, and assessing.

In recent years, there have been increasing voices that support the use of other terms to refer to the different functions of assessment. For instance, assessment *of* learning and assessment *for* learning are two terms that have gained considerable attention in the assessment field. Assessment *of* learning is used to denote summative assessment, while assessment *for* learning intends to connect assessment directly to the learning process (Baird et al., 2017; Bennett, 2011). According to Black and Wiliam (2018), assessment *for* learning refers to teachers using evidence drawn from assessment to inform their teaching, while assessment *of* learning refers to using assessment to decide whether students have achieved the learning objectives. Baird et al. (2017) further commented that assessment *for* learning distinguishes itself from formative assessment, as assessment *for* learning focuses more on student autonomy as an informal and ongoing process, while in some historical interpretations, formative assessment is more associated with teachers using test results as feedback.

However, there has also been criticism about this substitution. According to Bennett (2011), this change of terms could potentially result in oversimplification. He argued that the substitution removes the responsibility of summative assessment for supporting learning and further leads to oversimplifying the complex relationship between assessment and learning. Black and Wiliam (2018, p. 553) also commented that "these various 'prepositional permutations'" did not refer to the involvement of peers in supporting learning, and these changes in wording have made their meanings difficult to understand.

In the field of language education, terms such as dynamic assessment and learning-oriented assessment have also been proposed to reflect the intention to change the traditional relationship of assessment to learning. Grounded in Vygotsky's understanding of ZPD, dynamic assessment entails mediating learners' psychological process through using questions, models, prompts, and feedback while learners are having difficulties in completing the assessment tasks (Poehner & Lantolf, 2023). Instead of seeking to improve students' test performance, dynamic assessment aims at assessing whether learners can recognise and correct problems in their language performance, what assistance learners require to move forward, and what improvement occurs in learners' L2 performance (Poehner, 2008). Learning-oriented assessment, another concept identifying the concern of promoting better learning in the language education sector, seeks to bring learning, school-based assessment and large-scale assessment into alignment (Saville, 2021). Researchers advocating this concept tend to consider how diverse sources of evidence from assessment can complement one another and be utilised more productively (Jones & Saville, 2016).

For this study, the term 'formative assessment' will be used consistently throughout data generation, data analysis and discussion. The study adopts the definition of formative assessment from the Assessment Reform Group (2002, p. 1) as "the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there". There are two reasons for choosing this term. The first one is that in the context where this study is situated, the term 'formative assessment' and its Chinese counterpart "形成性评价" are widely adopted in literature and policy papers, and the use of this term can avoid misunderstanding to the greatest extent. The second and more essential one is that the focus of the term 'formative assessment' matches the focus of this study. According to Bennett (2011, p. 7), formative assessment is the "thoughtful integration of process and purposefully designed methodology

or instrumentation." It encompasses a wide range of teacher and student outcomes, including teacher eliciting evidence of students' learning, sharing learning expectations, providing feedback to move learning forward, students' supporting each other through assessment and feedback, and students' taking responsibility of their own learning. It does not absolve summative assessment from the responsibilities for supporting learning. On the other hand, assessment for learning and dynamic assessment place a stronger focus on the process of teacher using assessment to inform teaching instead of the roles summative assessment and peers could play in learning (Bennett, 2011; Black & Wiliam, 2018), while learning-oriented assessment has a more general focus on the overall assessment system, including classroom assessment, school-based assessment, and large-scale assessment. This study intends to explore teachers' classroom assessment activities, with a particular focus on how teachers adopted different activities to fulfil their various purposes. Formative, summative, self- and peer assessment at the classroom level are the focus of this study. Therefore, 'formative assessment' is a more appropriate key term for this study.

Many studies have been conducted on teachers' assessment methods in language classrooms and how they function and fulfil teachers' purposes. In her study, Rea-Dickins (2001) reported on classroom-based assessment practices in English classrooms in the UK. Two language-support teachers, one mainstream class teacher, and two learners from each of four classes were involved in the study. Using data such as classroom observations, lesson transcripts, and teacher interviews, Rea-Dickins (2001) identified three types of assessment used in EAL (English as an additional language) classrooms. The first one resembled the formal language tests taken by the learners. These helped teachers locate learners' achievement levels, make summative judgements regarding the curriculum goals, and provide instructions for individual learners for diagnostic and formative purposes. The second one happened within the informal interactions between teachers and the whole class. This type of assessment was performed as a part of the teaching activity. The teacher gathered evidence of students' learning by questioning and provided scaffolding and feedback during teaching. The third one took place within the small group work of learners. It was not pre-identified by the teachers in the interviews but was later observed in the classrooms. In the group work, the learners collaborated with their peers to work on phonics and words. The learners were observed self-correcting and peer-monitoring throughout the activity, which indicated that they took the opportunity to assess themselves and others.

These assessment types included formal ones and informal ones and involved both teacher and students as the initiators of the assessment activities.

According to the findings, Rea-Dickins (2001) argued that classroom-based assessment had three purposes. The first one is the bureaucratic purpose, as schools are responsible for sharing information relating to students' achievement with external agencies and other stakeholders. This purpose explains why assessment in language classrooms would share similar characteristics with formal language tests. The second one is the pedagogical purpose, which is an internal identity that meets the teachers' need for improving teaching and enhancing students' achievement. The third one is the learning purpose, which focuses more on the role of learners in the assessment process. According to Rea-Dickins (2001), this purpose motivates learners to participate in interactions and contributes to learning instead of measuring learning. The purposes of classroom-based assessment in this study appeared to have been influenced by different educational stakeholders, and formative purpose only guided a part of the observed assessment activities. The study highlighted several potential assessment strategies that teachers could draw on. It also conducted further analysis to address the purposes embedded within the strategies. However, the study did not discuss the balance between the assessment practices influenced by these purposes or whether a balance could be achieved. On the one hand, the bureaucratic purpose of formal assessment could not be denied for its importance in monitoring. On the other hand, the efforts made to promote learning in informal assessment are equally indispensable. Attention should, therefore, be drawn to the intricacies of classroom assessment in terms of how different purposes of assessment work together to create learning opportunities.

In their study, Hill and McNamara (2012) proposed a research framework for understanding classroom-based assessment based on an empirical study of two Australian school classes studying Indonesian as a foreign language. Through participant observation, interviews, and a grounded approach to analysis, they answered four questions which they considered essential in understanding classroom-based assessment: "What do language teachers do when they carry out classroom-based assessment? What do they look for when they are assessing learners? What theory or 'standards' do they use? Do learners share the same understandings? (Hill & McNamara, 2012, p. 398)" Particularly, three critical dimensions of assessment were attended to while answering these questions. The first dimension was evidence, which referred to the assessed construct, the assessment approach, and, the target

and agent of assessment (who is assessed by whom). The second dimension was interpretation, which referred to the criteria, theories and beliefs that guided the assessment practices. The third dimension was use, which referred to the purposes of the assessment.

Hill and McNamara's (2012) framework explores not only the different stages of teachers' assessment practices but also investigates their understanding of the subject knowledge and their underlying values regarding learning, teaching and assessment. By addressing these issues, classroom assessment activities are viewed as closely linked to the learning and instructional processes and serve as a facilitator alongside the learning process, not an after-learning event, which are essential features of assessment activities informed by socio-cultural learning theories (James, 2008). Especially, it pays attention to the learners' beliefs about language learning and assessment, which implies that learners and learning remain the focus of assessment. This framework can be particularly helpful in assisting researchers to understand the classroom-based assessment activities observed as it draws researchers' attention to the detailed elements within teachers' classroom assessment practices and the relationship between teachers' assessment practices and their understanding of learning, teaching and assessment. It views assessment as situated in learning and considers assessment as a social practice among teachers and students rather than an individual practice in nature, and, as a result, could serve as a theoretical lens for my classroom-based empirical investigations.

In the Chinese context, studies have been conducted to explore how assessment can be formatively used to improve students' learning. In a special issue of the Language Assessment Quarterly journal guest-edited by Poehner and Lantolf (2023), a number of papers were published to report empirical studies about the implementation of dynamic assessment in the Chinese L2 setting. The papers addressed several topics that have emerged in L2 dynamic assessment research in China, which included the scalability of the assessment procedures, the learning potential of dynamic assessment, the interpretation of learner abilities through dynamic assessment, and the importance of extending dynamic assessment research to additional language constructs and cultural contexts (Poehner & Lantolf, 2023). One of the papers, namely Zhang and Xi (2023), offered insights into my area of investigation, which refers to English language teaching and assessment in Chinese secondary classrooms. They conducted an experimental study on two secondary writing classes (one control class and one experimental class) to discover whether and how dynamic

assessment can facilitate the development of students' writing performance and metacognitive competence. The two classes were taught by the same teacher who was experienced in assessment research and integrating dynamic assessment into English language classrooms. Both classes received instructions regarding process and genre-based writing based on the same textbook and in-and-after class written tasks, with the experimental class being given more metacognitive instructions during the students' pre-writing, while-writing and post-writing activities.

With ratings of students' writing, self-report questionnaire data from the two classes (sixty-four students in total), and interviews with 8 stratified sampled students (4 higher attainment students and 4 lower attainment students), the authors found that students in the experimental class achieved considerably greater progress in English writing than those in the control class and have developed conscious awareness of metacognition. The students were also reported to have perceived the metacognitive instructions during dynamic assessment as changing their attitudes towards and building their confidence in English writing, suggesting that learner agency could be supported through guided activities that introduced writing strategies and offered opportunities to apply them. One limitation of this study is, however, that no classroom observations were conducted to observe how exactly the teachers implemented dynamic assessment in these classes and how teachers and students interacted with each other. To better understand the implementation and impact of formative assessment practices in English language classrooms in the Chinese context, more attention should be focused on the teachers' actual classroom practices and the teacher-student interactions during the assessment activities.

In this section, I have reviewed the changes in thinking and debates about classroom language assessment. Influenced by the sociocultural theory, classroom language assessment takes a role that is different from a language test. Not only does it focus on learning outcomes, classroom language assessment also pays attention to the learning process and how appropriate feedback can be provided to facilitate learning. However, it should be noted that few studies examined teachers' observed assessment practices against their espoused assessment beliefs, and vice versa. In the next part, the focus of this chapter will be shifted to teachers' enactment of language assessment policies. Educational policies regarding language teaching and assessment could be planned and released effectively, but implementing such policies is a more challenging task. The dimensions that could influence

the assessment policy enactments and the agency of language teachers will be addressed. Studies on the relationship between language assessment policies and teachers' enactment in classrooms will also be discussed.

2.3 Teacher enactment of language assessment policy

Besides changes in thinking about language learning and assessment, language assessment policies also have an impact on teachers' uses of language assessment in school systems (Cumming, 2009). Ideally, language curriculum policies, language tests, and classroom assessment of language teachers can work together to build a communicative learning environment and a formative assessment environment: curriculum policies present the performance descriptions for different language levels and emphasise the importance of students' communicative competence; the alignment between curriculum and assessment is highlighted in test design; the formative use of assessment is ensured in classroom teaching (Cumming, 2009). However, while educational policies could be planned and released effectively, implementing a policy is much more challenging. In most cases, teachers could have little say in the curriculum design or the educational policies they are involved in. While at the same time, teachers are not naïve actors – they could be autonomous practitioners, whose practices are affected by their own interests, values, and external translation and interpretation of the policies and may not entirely match what the policies expect (Ball et al., 2011). Such circumstances make teacher enactment of policies worthy of investigation, particularly when educational policy changes are introduced from the top to local classrooms without knowing whether and how teachers accept and/or adopt the changes.

Being encoded in texts and documents, a policy could be decoded in complex ways (Braun et al., 2011). Instead of being used as an unproblematic solution to a problem, a policy would be contested by parties from different backgrounds and interpreted according to their own values and needs. This process leads to the recontextualisation of the general policy ideas and results in contextualised practices – the final expression of policy enactment (Braun et al., 2011). According to Braun et al. (2011), policy enactment in education can be influenced by four contextual dimensions. The first one is the situated context, which refers to the location and history of the school settings that contribute to the translation and interpretation of policy from a broader perspective. The second one is the professional context, which relates to teachers' values, experience, and dedication regarding the policy, and the policy

management within schools. The third one is the material context, which refers to the physical aspects of schools, including available technology, the level of staffing, the surrounding infrastructure and budgets. The final one is the external context, which is concerned with the possible pressures and expectations from local authorities and other stakeholders. Working together, all these four dimensions create specific and dynamic contexts in different school settings. They are pertinent to language teaching and assessment practices and may result in various policy decisions and enactments in classrooms of different contexts.

Similar to educational policies in general, language assessment policies could be enacted in various ways under the influence of the four contextual dimensions. Firstly, the situated context could make the policy enactment more complicated if the assessment policy is adopted from a context with a different sociocultural tradition (Chen et al., 2014). Secondly, educational professionals, especially teachers and school administrators, could determine how the assessment policies are enacted in language classrooms, as they play a key role in translating the new ideas into reality (Gu, 2014). Thirdly, the physical aspects of schools, such as classroom layout and available technology, could restrain certain assessment practices. For example, classroom size and facilities could influence teachers' group or individual assessment management. Fourthly, the external context could also greatly influence the enactment of assessment policies, especially when teachers experience test pressure and have little choice but to focus more on improving test performance.

Other researchers explained teachers' roles as meaning-makers of policies and agents of change from a teacher agency perspective. Agency is a term that receives extensive attention in social science. In social theory, agency is often defined as "the capacity for autonomous social action" or "the ability to operate independently of determining constraints of social structure" (Biesta & Tedder, 2007, p. 135). It describes the ability to control one's actions or respond to a set of circumstances. Building on pragmatism, Emirbayer and Mische (1998) offer a three-dimensional way to understand agency, which, from their perspective, illustrates the complexity of the concept. The three dimensions are iterational, projective, and practical-evaluative dimensions, which represent "the influences from the past, the orientation towards the future and the engagement with the here and now" (Priestley et al., 2015, p. 23). These dimensions were further adopted by Priestley et al. (2015) in constructing

their teacher agency model, which can be used to explore teachers' agentic moves towards and interactions with assessment policies.

In this model, the iterational dimension refers to the influence of histories of a teacher, which include both general life histories and professional histories. According to Emirbayer and Mische (1998), the iterational dimension refers to:

> the selective reactivation by actors of past patterns of thought and action, routinely incorporated in practical activity, thereby giving stability and order to social universes and helping to sustain identities, interactions, and institutions over time. (p. 971)

Teachers can draw on many experiences from the past while designing and enacting assessment practices. For example, their professional education experiences could equip them with assessment theories, skills, and subject knowledge; their past working experiences in school could expose them to the assessment practices of other colleagues; their own learning experiences could provide them with assessment examples from their teachers. Forsberg and Wermke (2012) reported in their study that German and Swedish teachers considered their learning experience and colleagues' support valuable sources of assessment knowledge. Carless (2005) also noted that failing to build on past experience has contributed negatively to the assessment reform in Hong Kong. Thus, to understand teachers' assessment practices, it is important to probe their past experiences and explore the origin of their beliefs and what histories contribute to their agency in assessment practice.

The projective dimension looks at the teacher's short-term and long-term aspirations about their work, which guide the teacher's future actions. According to Emirbayer and Mische (1998), the iterational dimension refers to:

> the imaginative generation by actors of possible future trajectories of action, in which received structures of thought and action may be creatively reconfigured in relation to actors' hopes, fears, and desires for the future. (p. 971)

This dimension explains how teachers' long-term and short-term aspirations affect their assessment practices. On the one hand, such aspirations could be a product of teachers' prior experience, educational values, and beliefs, aiming for students' development and welfare (Lasky, 2005). On the other hand, teachers' aspirations could be more narrowly instrumental, as they might be shaped by the fabrication of school image and the performativity goals in

school settings (Ball, 2003). For Priestley et al. (2015, p. 105), performativity is a "demand on schools and teachers to 'perform', that is, to generate achievements in a clearly specified range of 'outcomes'." Such a demand might strongly influence teachers' assessment practices and conflict with their assessment values and beliefs (McMillan, 2003). Levy-Vered and Alhija (2015) also pointed out that if the accountability purpose of assessment is given priority in policy and practice, teachers' conceptions of assessment tend to respond negatively to such pressure.

While the other two dimensions are associated with the past and the future, the practical-evaluative dimension represents the influence of the teacher's day-to-day working environment, including the practical conditions in the context. According to Emirbayer and Mische (1998), the practical-evaluative dimension entails:

> the capacity of actors to make practical and normative judgements among alternative possible trajectories of action, in response to the emerging demands, dilemmas, and ambiguities of presently evolving situations. (p. 971)

For this dimension, Priestley et al. (2015) identify three aspects that contribute to the conditions through which teachers achieve their agency. The first is the cultural aspect, which refers to the culture of the situated contexts, such as the classes, the schools, and the wider society. Fleer (2015), for example, identified the tensions that emerged while teachers worked against the discourse of traditional test-oriented assessment culture. The second is the structural aspect, which refers to the social and power relationships in a teacher's workplace. School-level policies on assessment, school managers' support for assessment, and expectations and demands from parents and the local community can influence teachers' assessment practice (Liu & Xu, 2017). The third is the material aspect, which refers to the resources and the physical environment that encourage or impede teachers' agency. As Xu and Harfitt (2019) suggested, contextual factors such as large class size, limited opportunities for individual feedback, and immense teaching and marking responsibilities could determine whether a form of assessment practice could be successfully conducted.

Together, the three dimensions enable and constrain teachers' capacity for and achievement of agency, and shape their assessment practices in an ongoing ecology. These dimensions help understand teachers' assessment practice and explore how teachers interact with the past, present, and future. Agentic teachers can identify opportunities to implement formative

assessment and exert their professionalism. Their practices are underpinned by personal dispositions and the objective contexts in which their practices are enacted (Molla & Nolan, 2020).

Although there are numerous studies conducted on policy enactment in the classroom language assessment field, I will closely look at two studies associated with the Chinese context. The first study, conducted by Gu (2014), is about curriculum and assessment policy implementation in the Chinese mainland. According to Leong et al.'s (2018) review of formative assessment research that has been conducted in East Asia and my own search on Chinese databases for digital publication resources, this is the most recent and detailed empirical case study about formative assessment policy implementation conducted in a secondary English language classroom in the Chinese mainland. A research gap is thus identified – there seems to be a severe lack of knowledge of how English language teachers implement formative assessment and respond to the current formative assessment policy in actual classroom settings in the Chinese mainland context. In the study, Gu (2014) investigated classroom assessment practices under the newly implemented English curriculum in a key secondary school in Beijing. The participant was an English teacher with ten years of teaching experience. An interview with the teacher and observations in her class was used to collect data. In the interview, Gu (2014) tried to elicit the teacher's understanding and comments on the concept of formative assessment, as this concept has been written in the curriculum documents for teachers to apply. The teacher was not sure of the meaning of the concept. She also claimed in the interview that it was not required for teachers in this school to read the new curriculum documents unless they needed to "participate in teaching competitions at district, municipal or national levels" (p. 292). In the classroom observation, Gu (2014) found that the feedback types the teacher used in class were primarily evaluative and descriptive. The former indicated right or wrong, and the latter related to vocabulary and grammar explanations, as these were the things that would be tested.

Although there was only one participant in Gu's study, the findings still suggested that the curriculum and assessment policy remains a "documentary rhetoric" in this case (Gu, 2014, p. 298). Multiple reasons could contribute to the described practice. The first would be the high-stake examinations such as Gaokao, which shape the classroom's external and internal educational context (Gu, 2014). As Gu (2014) reported, the assessment types the class adopted were mainly tests conducted at the end of a unit, semester, or year. Most of these

tests imitated the format of English Gaokao, and rankings were done at both class and school levels. This reflected the external context dimension, which could have an extensive impact on the enactment of assessment policies, and the projective dimension, which is associated with what the teacher and the school expected of the students. Although the daily tests merely imitated the English Gaokao, they were still as powerful as the real Gaokao because ranking systems were introduced. Students and teachers were evaluated by test performance, which would drive them to work harder for higher scores. The ranking itself would also cause their anxiety about being exceeded, further resulting in their overemphasis on tests and downplaying the significance of other means of assessment.

The second reason would be the teacher's understanding of the newly implemented English curriculum and suggested assessment methods (Gu, 2014). According to Gu (2014), the participant showed limited knowledge regarding the concept of formative assessment. She understood assessment as equal to testing, "not as part of teaching and learning" (p. 297). This reflected the professional context dimension, which directly impacted how assessment practices were conducted, and the iterational dimension, which referred to the teachers' prior knowledge and professional learning experiences regarding assessment. As the participant reported, she and her colleagues did not receive any training in assessment and were not confident about designing their own assessment in the classroom. She also considered assessment to be the job of experts, not the job of teachers. If teachers are not able or confident to implement the assessment practices as instructed, or if they do not even agree with the vision of policymakers, it could be imagined that teacher enactment of such policies would encounter difficulties. Changes would be hard to take place if these educational professionals' beliefs and values were not changed.

The third reason would be associated with the school management level (Gu, 2014). According to the participant, the school did not require teachers to read the new curriculum documents. The only events requiring their knowledge of the curriculum were the teaching competitions held at the regional or national level, which in this case seemed more like a show instead of an assessment or an incentive for teachers. This could also reflect the professional context dimension and the practical-evaluative dimension and partially explain the second reason, which is associated with the teacher's neglect of the new curriculum and assessment policies. The school management level provided little incentive for teachers to learn new educational ideas, such as formative assessment, which therefore remained a

vague concept for teachers. The possible reason for this could be attributed to the fact that the school was also under accountability pressure. If teachers' current practices could guarantee students' test performance, there would be little reason for the school to take the risk of promoting new ideas and practices. The different contextual dimensions interacted with the school setting and resulted in the practices observed in Gu's (2014) study. Under the influence of many factors, the policy is hovering above the field instead of touching the classroom ground.

Although this study expressed concerns over the implementation of formative assessment in a test-dominated context like the Chinese mainland, Gu (2014) has expressed interest in other attempts to implement more balanced assessment systems in similar contexts, including efforts made by educational professionals in Hong Kong, a Special Administrative Region (SAR) in China. In another article, Carless and Lam (2014) discussed the potentials and challenges of implementing formative assessment in Hong Kong and argued that the assessment practice that can be accepted by the local practitioners is the one that does not conflict with the dominant testing paradigm. They reported two cases of English classroom assessment practices in Hong Kong primary schools that reflected their thinking on combining summative and formative assessment. In the first case, the teacher encouraged students to generate their own mock tests, led them to a more meaningful preparation of the actual test, and helped them obtain some ownership of the test-preparation process. After revising some key content of the test, the teacher asked the class to form into four groups and develop a mock exam of their own. During this process, the teacher monitored their work and provided feedback when needed. The groups exchanged their exam papers after completing their mock paper and were responsible for marking their peer's papers. In the end, the groups would present the exam they constructed, and the teacher would then give feedback to clarify any misunderstanding regarding the test items. The teacher reported that the overall grades of his classes improved after implementing this practice. However, limited evidence showed whether this approach contributed to students' achievement. Some students also argued that it was difficult to develop a test due to their limited command of English, while some others reported that not every student was actively engaged in the construction of the mock exams.

The second case was about peer cooperation that focused more on test follow-up, which falls within the notion of using test data formatively. In this case, the teacher intended to actively

involve students in the post-test follow-up to avoid the conventional methods of teacher explanation and input. She adopted two methods. The first was a student presentation of test strategy, which involved high-achieving students sharing how they reached the correct answers. The second was conducting a group activity in which students developed a 'perfect correction sheet' for the test. Students were encouraged to collaborate in a team and pool their ideas while developing the sheet. The teacher reported the two strategies as effective because they allowed low-achieving students to learn from high-achieving students. However, she also acknowledged that those less-confident students might be left out of the peer cooperation process, as high-achieving students may dominate the whole process.

Generally speaking, Carless and Lam (2014) were very positive about this kind of combination of summative and formative assessment. They argued that by using these formative strategies, students could appreciate that tests do not have to be only about scores and performance. The two participants were also found to be impacted by their professional learning experiences. They have both attended teacher training courses on assessment for learning taught by the two authors. They both understood the concept of formative assessment and have attempted to apply their understanding of this concept in their classroom practices. However, the study did not report whether the teachers used these assessment approaches regularly or whether it was a one-off practice for the research. It should also be noted that both examples were taken from primary school classrooms, where exam pressure is less intense. Such practices were acceptable in primary school settings, but whether it is feasible in middle school or high school settings is yet to be learnt.

Lam (2016) and Qian (2014) stated that different contextual dimensions restricted formative assessment practice in Hong Kong. Despite the widely implemented assessment reform, formative assessment has become a controversial topic in Hong Kong, as learning is still test-driven (Lam, 2016). Berry (2011a) reported that many teachers had expressed concerns over reliability, validity, and fairness issues about formative assessment and complained about not receiving enough training and support. Although students were reported to be receptive to the new assessment methods, teachers are reported to have experienced many challenges, including an excessive workload for students, failure to integrate formative assessment with the regular English curriculum, and the lack of transparent communication among teachers, parents and school administrators (Lam, 2016; Qian, 2014). Carless and Lam (2014) have reported two inspiring cases and have suggested a way of implementation

that could help integrate formative assessment into a test-dominated context. Although the contextual dimensions remain to be powerful constraints in the context of their study, which suggests that there is still a long way to go, it could not be denied that given adequate training and enough space of freedom, teachers would be able to innovate in assessment practices in the testing-dominant paradigm.

2.4 Summary

In this chapter, I have reviewed three sets of studies that are related to the research focus of this study. The first set is concerned with the development of language theories. From behaviourism to sociocultural theory, the focus of language teaching and learning has moved from individual habit formation to communication and then to collaboration with peers and experts. The second set is concerned with the development of language assessment. Informed by different educational and language theories, the focus of language assessment has moved from discrete linguistic elements to integrative language skills. The purpose of language assessment is no longer limited to measuring language ability but also includes a focus on the language learning process. The third set is concerned with teacher enactment of language assessment policy. Changes in thinking of language theories and language assessment can lead to new designs of language assessment policies, but it is through teacher enactment that these policies impact the classroom. With different dimensions in the context and within teacher agency, teacher enactment may vary in different settings, and actual language teaching and assessment activities in classrooms may not resemble what has been promoted in curriculum and assessment policies. The existing literature will be borne in mind throughout the design of the methodology, which will be addressed in the next chapter, and the analysis of the generated data in the following chapters.

Chapter 3 Methodology

The research problem of this study focuses on the enactment of formative assessment and the possible mismatch between assessment practices, the teachers' understanding, and the assessment policy regarding English teaching and assessment in China. In order to understand the problem and explore the reasons behind assessment practices, three sets of questions are proposed:

> RQ1: What classroom assessment activities do teachers adopt? Why do they choose these activities?
>
> RQ2: What are teachers' understandings of assessment, in particular formative assessment? Why do they adopt such beliefs?
>
> RQ3: How is the current assessment policy reflected in the textbook, teacher guidebook and the Zhongkao teachers work with? What actions have the teachers taken to put the assessment policy into practice under the influence of these materials? Why do they enact the assessment policy as such?

The research questions cover the three perspectives that are most relevant to the research problem and reflect how the study is situated in larger disciplinary issues in social science research regarding the ontology and epistemology of social phenomenon. Ontology is "the study of being" (Crotty, 1998, p. 10). It concerns the nature of being and refers to "what exists in the human world that researchers can acquire knowledge about" (Moon & Blackman, 2014, p. 1167). In contrast, epistemology concerns "the question of what is (or should be) regarded as acceptable knowledge in a discipline" (Bryman, 2016, p. 24). It refers to a way of understanding the questions of "in a world where all sorts of knowledge exist, how do we know which to trust, which are meaningful" (Braun & Clarke, 2013, p. 28) and deals with "the nature of knowledge, its possibility, scope and general basis" (Hamlyn, 2005, p. 260). According to Crotty (1998, p. 10), it is difficult to discuss ontology and epistemology separately, as "to talk of the construction of meaning is to talk of the construction of meaningful reality." At the same time, the philosophical perspective a researcher adopts can fundamentally affect his or her views of the world, which will further guide the theoretical perspectives, methodology, and methods he or she chooses (Waring, 2017). As a result, this

chapter aims to outline how this study is situated ontologically and epistemologically (section 3.1), and the adopted theoretical perspectives and methodology (section 3.2) which are aligned and informed by the ontological and epistemological position. Research design (section 3.3) informed by these positions is introduced in detail, and ethical considerations (section 3.4) are also addressed.

3.1 Ontology and epistemology

There are two poles in ontological positions – realism and relativism. Realist ontology believes that one single truth exists, and the truth exists independent of human experience (Bryman, 2016; Moses & Knutsen, 2012). On the other hand, relativist ontology believes that no one truth exists. Instead, reality is constructed in people's minds and reflects different individuals' differing experiences and perspectives (Braun & Clarke, 2013).

Between these polarised ontological positions, the degree of confidence in defining the nature of truth varies, which leads to different categories of realism and relativism. For example, while naïve realism claims that one true reality exists and is driven by "immutable natural laws and mechanisms" (Guba & Lincoln, 1994, p. 109), structural realism accepts that although one true reality can be identified, the definition of the reality "can change, at which point the nature of reality also changes" (Moon & Blackman, 2014, p. 1170). Another realist approach, critical realism, assumes that "there is a world of events out there that is observable and independent of human thoughts" but argues that "knowledge about this world is socially constructed" (Denzin & Lincoln, 2005, p. 13). On the other side of the ontological spectrum, different types of relativism also view the nature of reality differently. Although relativists generally argue that realities are "local and specific constructed and co-constructed" (Guba & Lincoln, 2005, p. 193), there are also different beliefs about who constructs the realities and whether they share the realities. For example, Moon and Blackman (2014, p. 1170) propose the concept of bounded relativism, which argues that "one shared reality exists within a bounded group, but across groups different realities exist." The groups can refer to different cultures or different moral groups (e.g., anthropocentrism and ecocentrism). On the other hand, relativism argues that a shared reality does not exist. Rather, each individual constructs his or her own reality.

By addressing different ontological positions, it is evident that there can be different and even conflicting ideas of reality. For this study, I will adopt bounded relativism as the ontological position. The reason is that as a bounded group, language teachers in China can be expected to construct a different reality around formative assessment practices and values from those of their counterparts in other parts of the world. As I have argued in Chapter 1, teachers' assessment practices and values can be influenced by their prior educational and work experience, the power relationships in their workplace, and the contexts where assessment takes place. Language teachers in China may understand the concept of formative assessment or adopt the idea differently from the originally intended concept or related practices. On the one hand, there may be some shared reality in the context where formative assessment was brought up and developed; on the other hand, there may be a different reality in China that constructs a different version of the concept so that it could be acceptable and applicable. From this ontological position, what language teachers in China believe and what the Western concept of formative assessment intends or instructs are viewed as equally important. This ontological position provides the foundation for acquiring knowledge of Chinese teachers' understanding regarding assessment practice, values, and policy.

As previously identified, ontological and epistemological issues tend to emerge together rather than separately. An ontological position informs an epistemological position. Like the ontological spectrum, there is also a range of epistemologies. Firstly, there is objectivism. Objectivism is often considered to imply realism (Moon & Blackman, 2014), as objectivists believe that "social phenomena and their meanings have an existence that is independent of social actors" (Bryman, 2016, p. 29). On the other side of the spectrum, there is subjectivism. Subjectivism holds that knowledge is "always filtered through the lenses of language, gender, social class, race, and ethnicity" (Denzin & Lincoln, 2005, p. 21). For subjectivists, meaning is "imposed on the object by the subject" instead of emerging from the interaction between subject and object (Crotty, 1998, p. 9). Subjectivism is closely associated with ontological relativism, as both positions are person-centred and focus on how people's experience shapes their making sense of reality.

Another epistemological position that is more compatible with the chosen ontological position of this study is constructionism. Constructionism rejects the view that objective truth is waiting to be discovered. Instead, "what we know of the world, and ourselves and

other subjects in the world is constructed (produced) through various discourses and systems of meaning we all reside within" (Braun & Clarke, 2013, p. 30). Constructionists believe that meaning is constructed. Different people may construct meaning differently regarding the same phenomenon. How they construct meanings and understand the world depends on their social, cultural, and historical backgrounds (Bryman, 2016; Denzin & Lincoln, 2005). According to Moon and Blackman (2014), constructionist studies aim to generate a contextual understanding of a defined problem. In this vein, this study aims to understand language teachers' interpretation and enactment of formative assessment in the Chinese context. Does their educational and work experience affect their assessment practices? What factors constrain and enable their choices of assessment practice? Does the local conventional practice of assessment shape their attitudes towards formative assessment? Does the current education system affect their understanding and response to the current assessment policy? From these questions, the context is understood to play a key role in understanding and describing the reality. Consequently, bounded relativism and constructionism are the ontological and epistemological positions underpinning the study.

3.2 Theoretical perspective and methodology

The ontological and epistemological positions of a researcher will guide their philosophical perspective. This perspective will frame the researcher's theoretical perspective and methodological design. In this section, the theoretical perspective of the study, which is developed from the adopted ontological and epistemological positions, will be presented. The methodology informed by this theoretical perspective will also be outlined.

Crotty (1998, p. 66) argues that a theoretical perspective refers to "the philosophical stance lying behind a methodology." Put another way, it is a set of assumptions that drives the way research is conducted. Like ontology and epistemology, there is a range of possible theoretical perspectives. According to Moon and Blackman (2014), the spectrum of theoretical perspectives covers the range from deductive acquisition of knowledge to inductive acquisition of knowledge. On the deductive side, knowledge is more objective and generalisable, and experimental research design is an appropriate process for knowledge production. In such cases, knowledge is about "the relationship between interventions and the consequences of these interventions on a particular phenomenon" (Biesta, 2010, p. 101). Researchers committing to this perspective usually adopt the theoretical perspective of

positivism, which holds that objects have meanings independent of any consciousness of them, and knowledge can only be accurate if it is gained through objective application of scientific method (Hughes & Sharrock, 1997).

On the inductive side, knowledge is understood to be more value-laden and contextually unique. Naturalistic research design is the main example of knowledge acquisition under this perspective, and knowledge here is about the phenomenon observed. Researchers committing to this perspective may adopt interpretivism as their theoretical perspective. While the positivist approach seeks to identify one true knowledge, the interpretivist approach "looks for culturally derived and historically situated interpretations of the social life-world" (Crotty, 1998, p. 67). From an interpretivist point of view, to understand a social action, the interpreter should view the action as meaningful, show respect and fidelity to the actors' lifeworlds, and recognise the contribution of human subjectivity to knowledge and action (Schwandt, 2000). In order to achieve such an understanding, the interpreter should not only focus on the particular action but also seek to grasp the whole context in which the action takes place (Outhwaite, 1986). Since this study is a social science study taking a bounded relativist and constructionist position, the theoretical perspective of this study is more appropriately situated on the more inductive side of the spectrum, adopting an interpretivist perspective. Instead of depending purely on the realities observed, the study adopts a hybrid philosophical orientation combining both inductive and deductive knowledge acquisition, putting both theories and realities at the centre of analysis. The inductive approach is driven by the data, involving identifying the information pertinent to the research questions and encoding it before the interpretation process, while the deductive approach is driven by the theories reviewed, such as the different language theories, theories related to the relationship between language learning and assessment, the teacher agency model, and the contextual dimensions of policy enactment, which help to explore and tag the data (Braun et al., 2019).

With the broad theoretical perspective identified, I can now describe the methodology of the study. The study adopts a case study approach. According to Bakker (2012, p. 487), "case study research is often associated with an emphasis on the importance of interpretation of human meaning." By examining a case or multiple cases of the subject of the study, a researcher can explore the totality of a situation and produce detailed descriptions of social actions, which are central for interpretation. This study aims to understand teachers' enacted

assessment practices in the local context and their understanding of formative assessment and the current assessment policy. What teachers practise is observed in their classrooms; what they believe and understand about assessment approaches and policies are elicited through interviews and further explored through classroom observations. As a result, it is important for this study to look closely at teachers and to understand their work environment as a whole. A case study approach allows for detailed descriptions of teachers' actions and investigates what each teacher thinks. Together, these evidence sources enable me as the researcher to interpret the phenomenon of classroom assessment in the targeted context.

According to Stake (1994, p. 237), a case study is "both the process of learning about the case and the product of our learning." It is the study of the "particularity and complexity" of a single case or sets of cases, "coming to understand its activity within important circumstances" (Stake, 1995, p. xi). The definition of a 'case' remains subject to debate. As Smith (1978) and Stake (1994, 1995) argue, a case is understood as a bounded system. They argue that it is common to identify certain features within the boundaries of a case, and "the boundedness and the behaviour patterns of the system are key factors in understanding the case" (Stake, 1994, p. 237). Other researchers like Yin (2018) disagree with the 'boundary' argument. In his book, Yin (2018) addresses the scope of a case as there is no clear boundaries necessarily between a phenomenon and its context. He defines a case study as:

> A case study is an empirical method that investigates a contemporary phenomenon (the 'case') in depth and within its real-world context, especially when the boundaries between phenomenon and context may not be clearly evident. (Yin, 2018, p. 15)

In other words, if you aim to understand cases, such an understanding will require knowledge of pertinent contextual factors. In such circumstances, a clear distinction between the features of the case and its context is not necessary. As this study adopts bounded relativism as its ontology and is conducted within one cultural context, I would argue that shared context will be a significant factor that affects the shape of any reality of assessment practices. To understand the cases in this study, the relationship and interaction between the broader context and what goes on in each case should be kept in mind. On one level, the cases of this study appear to be bounded cases, as the teachers work in different settings (which will be further explained in section 3.3.3). However, on another level, they also share the same cultural and policy background, which makes any boundary between them rather blurry and

tightly interwoven with the context. Yin's approach to the case informs this study and will be considered throughout the research design in the next section.

3.3 Research design

According to Yin (2018, p. 27), the design of case study research should include five components: "a case study's questions, its propositions, its case(s), the logic linking the data to the propositions, and the criteria for interpreting the findings." To be more specific, the design should address the definition of research questions (section 3.3.1), the identification of study propositions (section 3.3.2), the selection of cases (section 3.3.3), the generation of data (section 3.3.4), and the analysis of data (section 3.3.5). These five parts will be discussed below.

3.3.1 Research questions

The study aims to explore teachers' activities and understandings of formative assessment and their enactment of assessment policies in English classrooms in the context of Chinese middle schools, and the research questions of this study can be classified into two kinds. The first kind of questions is descriptive and mainly asks 'what' and 'how'. For example: What classroom assessment activities do teachers adopt? What are teachers' understandings of assessment, in particular formative assessment? How is the current assessment policy reflected in the textbook, teacher guidebook and the Zhongkao teachers work with? What actions have the teachers taken to put the assessment policy into practice under the influence of these materials? Descriptive questions answer "what is happening?" in the targeted context (Yin, 2012, p. 5). They also offer rich descriptions and insights, which enable further review and analysis. The second kind of questions mainly asks 'why'. For example: Why do the teachers choose these activities? Why do they adopt such beliefs? Why do they enact the assessment policy as such? These explanatory questions answer "why something happened" (Yin, 2012, p. 5). They can help construct the mutually constitutive relationships of different factors within the cases and enrich the understanding of the addressed phenomenon "beyond what can be discerned by using experiments or quasi-experiments alone" (Yin, 2012, p. 89).

3.3.2 Study proposition

While research questions point to the key interests of the study, propositions point to what should be studied (Yin, 2018). According to Yin (2018, p. 27), propositions are the ideas that "direct attention to something that should be examined within the scope of study." They reflect the theoretical issues reviewed and help identify the cases and relevant information to be collected. In the research questions, three important theoretical themes relevant to the study are identified: assessment activities, teachers' understandings of assessment, and assessment policy enactment. As the key theme of the study, assessment activities will be given due attention. Assessment activities in the cases will be considered in terms of the design and the feedback involved. The nature of alignment between assessment, curriculum, and pedagogy will also be explored in each case. As teachers are the actors that determine the realisation of educational theories and policies in classrooms, teachers' understandings of assessment will be explored, and their purposes behind their activities will be given special attention. As the assessment policies and the learning contexts provide external influence on classroom practices, how teachers' interpretations of assessment policies and the wider context of education affect teaching, learning, and assessment will also be explored.

3.3.3 Case selection

After identifying the research questions and the study propositions, a case or cases should be defined and selected for investigation. As Yin (2012, 2018) argues, what constitutes a case needs to be defined as the unit of analysis in case study design. He also argues that research questions and study propositions are needed in this process in order to arrive at a definition. Since the research questions and study propositions are related to how teachers conduct classroom activities and reach their thinking in the local environment, namely the institutional context and the general assessment context, the 'case' for this study is defined as the selected schools, which involve the teacher participants and their respective classes so that teachers' practices and thinking can be understood within the policies and practices of the participant schools, which are local responses towards the current assessment policies. The cases were selected from middle schools in Shenzhen city of China. The reason for choosing this context was that Shenzhen was undergoing a Zhongkao reform while the study was conducted. As an important test in Chinese middle students' educational experiences, Zhongkao significantly impacts middle school teachers' assessment practices but has not

received adequate academic attention that matches its influence. The process of policy changes around assessment practices provided relevant conditions for the study. It allowed me as a researcher to understand teachers' actions and thinking under the background of changes and enabled the study to be timely and important.

Instead of looking at one single case, the study adopts a multiple-case study approach. The reason for focusing on multiple cases is that this approach could make the evidence of the study more robust. The availability of schools pragmatically constrained the number of cases. The data generation began in September 2020, when schools in Shenzhen were reopened from COVID restrictions. With the constrained availability of schools, I managed to gain access to two middle schools in one district of Shenzhen and stayed in each school for approximately two weeks. In each school, two teachers, one early-career teacher (one year of teaching experience) and one experienced teacher (more than ten years of teaching experience), were recruited from Grade 8. The reason for choosing Grade 8 experienced and early-career teachers as participants was that the study intends to explore the assessment activities, understandings, and enactment of teachers with distinct professional backgrounds, and within the two schools, Grade 8 teachers demonstrated their willingness to participate in the study first. The two schools and the four teachers were given codes according to the order I approached them for data generation. The first school I approached was coded as School A, and the experienced and early-career teachers in this school were coded as T1 and T2 respectively. The second school I visited was coded as School B, and the experienced and early-career teachers in this school were coded as T3 and T4. Figure 3-1 shows the general design for case selection.

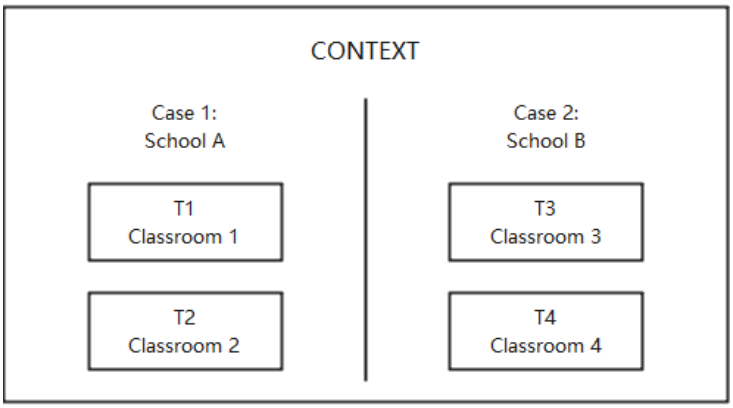


Figure 3-1 Design for case selection

3.3.4 Data generation

As Crotty (1998, p. 3) argues, methodology is the design that lies behind "the choice and use of particular methods and linking the choice and use of methods to the desired outcome." Case study research is not limited to a single method of data generation. It can draw on various methods, as long as they serve the research aims and design. As a result, the data generation methods chosen for this study serve the purpose of the research questions and aim to provide evidence for further data analysis.

Three modes of data generation were chosen for this study: non-participant observations, semi-structured interviews, and document analysis. The first method is non-participant observation in classrooms. Since a case study takes place in and focuses on a lived setting, non-participant observation of what is happening focuses on teacher-and-student interactions in classrooms, including teaching, learning, and, most importantly, moments or incidents of assessing learning. The classroom observations were intended to cover the processes of the teachers teaching a whole unit, as this allowed me to observe a whole curricular cycle, including both formative and summative assessment opportunities. A total of 32 lessons were observed (seven lessons for T1, eight for T2, eight for T3, and nine for T4). The classrooms observed were recorded by two audio devices. One audio device was placed on the teachers' desk at the front of the classroom, and the other was placed at the back of the classroom where I sat. This ensured that teachers' talking could be recorded with good quality even when they walked around the classrooms. An observation scheme was designed for the convenience of notetaking in class. It targeted four types of information: tracking time, the topic of the lesson, what action or activity is required, and what assessment opportunity emerges (see Appendix 1). The observed moments of assessment interactions were identified, transcribed and/or translated by me and used as an important data source for analysis. Since the purpose of this study is to elicit the meanings behind the data, the transcription intends to follow a "content-oriented analysis" (Nikander, 2008, p. 225) by involving the words the participants used, how they said their words, and any physical actions that happened at the same time in the transcripts if they were relevant to the analysis (Kowal & O'Connell, 2014).

The second method for data generation is semi-structured interview. While observations can help answer the 'what' and 'how' questions, semi-structured interviews can be constructive for answering the 'why' questions about teachers' understandings (Yin, 2018). Unlike survey

research, “in which exactly the same questions are asked to each individual,” semi-structured interviews create unique conversations between researchers and interviewees, “as researchers match their questions to what each interviewee knows and is willing to share” (Rubin & Rubin, 2005, p. 4). This method suits this study, as it adopts an interpretivist perspective and aims to understand the meaning made by the interviewees of what is happening. It can help me realise how the interviewees construct meanings within their context and understand the phenomenon of assessment practice of each case in depth. Two interviews were conducted with each of the teachers. A before-observation interview was conducted at the beginning of the unit’s teaching, and an after-observation interview was carried out at the end of the curricular cycle. Mandarin Chinese, the common language in this context, was used in the interviews to ensure that the meanings were clearly conveyed and received. An audio device recorded the interview processes, and the recordings were later transcribed, and translated, if necessary, by me for further analysis.

A semi-structured interview contains a number of predetermined themes and suggested topics, but there is also “openness to changes of sequence and forms of questions in order to follow up the specific answers given, and the stories told by the subjects” (Kvale, 2007, p. 52). This means that during a semi-structured interview, two jobs need to be done: the first is to address the prepared line of inquiry, which reflects the research interests, research questions, and study propositions; the second is to ask additional questions to pursue lines of inquiry as they emerge from the responses of the interviewees. The flexible format allows semi-structured interviews “to reveal how case study participants construct reality and think about situations, not just to provide the answers to a researcher’s specific questions and own implicit construction of reality” (Yin, 2012, p. 12). The before-observation interviews were intended to explore teachers’ background information, their practices and understandings regarding learning, teaching, and assessment in classrooms, their professional learning experiences, and their opinions towards the current assessment policies (see Appendix 2 for the interview schedule). The after-observation interviews drew on the observational data, enquired teachers about their motives behind their practices, sought further clarifications about their teaching objectives and reflections on assessment activities, and offered opportunities for teachers to clarify their previous statements (see Appendix 3 for the interview schedule for each teacher). For questions that emerged during the observations, I asked the teachers right after class to ensure that the context of the questions remained

familiar. Such informal interviews were recorded with an audio device, and notes were taken on the observation scheme to supplement the observational data afterwards.

The third method is document analysis. In this context, the textbook, teacher guidebook, and the Zhongkao test design teachers work with reflect the current assessment policy and have a direct impact on teachers' assessment activities. To illustrate the context under investigation, it is important to understand how these materials convey the assessment policy. These materials were therefore analysed to provide data on the context and corroborate evidence found in the classroom observations and the interviews (Bowen, 2009). The textbook and teacher guidebook were read through to explore the textbook design and any assessment-related suggestions. The Zhongkao test paper was also examined to understand the test design. Such an analysis served as the provider of background information on the situated context and offered a source of reference for further evaluation of the teachers' assessment practices, understandings, and enactment.

The three selected data generation methods provided data to address different parts of the research questions. The classroom observations helped answer a part of the descriptive research questions about what assessment activities happened in these classrooms and provided topics for further interviews and analysis. The interviews targeted the explanatory research questions and captured interviewees' understanding of assessment ideas and assessment policies. The document analysis provided background information and references for understanding the context. This layered design allowed me to identify the recurring themes related to the research problem and understand how the teachers constructed meanings in their context.

There are, however, limitations to these three methods. For document analysis, according to Bowen (2009), retrieving documents can be problematic when access to documents is blocked. It is possible that specific documents relevant to the teachers' teaching and assessment are not made public. Approaching teachers as insiders for help becomes crucial. For non-participant observations, as Liu and Maitlis (2010) suggest, one of the biggest challenges for non-participant observations is the observer effect, which may produce reactivity and self-consciousness in participants under study. Williams (2008) also argues that while non-participant observations typically involve strategies such as field notes and video/audio recordings, participants may interpret these strategies as intrusive, which might

potentially influence their behaviours. During the data generation, such effects were mitigated through longer periods of observation and building rapport with teachers, telling them that I was a learner learning from their teaching, not a supervisor or inspector.

Limitations also exist for semi-structured interviews. Opdenakker (2006) argues that, like non-participant observations, interviews may influence interviewees' behaviours. Such an interviewer effect may come from the interviewer's intentional or unintentional guidance in the interview. This limitation can be diminished by raising the interviewer's awareness of such an effect, and interview questions should be carefully chosen and reviewed so that interviewees will not be guided by these questions to make ungenuine responses. In addition, the interviewer effect may also stem from the power relationship between interviewers and interviewees, as Karnieli-Miller et al. (2009) and Barlow (2010) mention. The power relationship between interviewer and interviewees can be affected by many factors, including the choice of words in interviews, the content of inquiry, and the context in which the interview is carried out (Karnieli-Miller et al., 2009). Such a relationship may result in risks of discomfort and inconvenience for the interviewees and may further affect their responses in the interview. To mitigate any possible distress, the interviewer needs to make careful choices of wording. As Kvale (1996, p. 33) suggests, researchers should approach interviewees with "deliberate naïveté", which indicates an openness to any new and unexpected opinions expressed by the interviewees. Theoretical terms should be avoided in interviews, and the terms chosen by the interviewees should be reflected as much as possible (Kvale, 1996). The power can also be shared with the interviewee by sharing transcripts with the interviewees and providing opportunities for changes to any comment they later regret (Rowlands, 2021).

The other challenge for the three methods and the study more broadly is that two languages, Chinese and English, were involved in the data generation, analysis, and report. The researcher is not a trained translator and might not deliver all the meanings behind the data. However, several approaches are taken to minimise the negative impact. On the one hand, while translating texts from the document materials and transcripts from interviews and observations, the meanings behind the data are kept in mind, and word-for-word translation is avoided if such a translation might impede readers' comprehension. On the other hand, translation of key terms adopted different methods. First, the terms that exist in both Chinese and English academic literature on assessment, for example, 'formative assessment' and its

Chinese counterpart “形成性评价”, were used directly in data generation and thesis writing. Second, some terms that emerged in the data do not exist in English academic literature, for example, the term “过程性评价” (first mentioned and explained in section 5.4.2, pp.112-113). For this kind of terms, a word-for-word translation method was implemented to keep the original meaning (for instance, 过程性评价 – process assessment), and further explanations about the intended meaning behind the terms were provided to assist readers’ comprehension.

3.3.5 Data analysis

Since the study is a multiple-case study within a situated context, the data analysis begins by analysing the context based on the relevant document materials and then analysing each case, which leads to interpreting the findings in similarities and differences across the two cases. In this way, “the integrity of an entire case” can be retained, and “any within-case patterns across the cases” can be compared or synthesised within the situated context (Yin, 2018, p. 196). The documents were analysed with a content analysis approach, which organised information relevant to the research problems into categories (Bowen, 2009). The documents were skimmed, read, and interpreted for their underlying assessment ideas and were cross-checked to see whether these ideas were consistent with one another. Questions regarding whether and how teachers acted according to these materials were proposed for further observational and interview data analysis.

Then, within-case analysis and cross-case synthesis were conducted (Huberman & Miles, 1998; Yin, 2012, 2018). According to Huberman and Miles (1998), within-case analysis usually deals with two levels of understanding. The first level is descriptive, answering the questions of ‘what’ is happening and ‘how’ things are happening. “To tell a story” with the details and evidence observed is a good way of describing the “local actors, events, and settings” of each case (Huberman & Miles, 1998, p. 188). The second level is explanatory, answering the ‘why’ questions. Such explanations can serve multiple purposes, such as providing information, giving reasons, or supporting claims. Classroom observations and semi-structured interviews assisted these two levels of understanding, as outlined below.

Classroom observations served for the first level of understanding because this method provided empirical evidence that helped recognise observable aspects of assessment

activities as they occurred in classrooms. Teachers' assessment activities were first identified and highlighted on the observation schemes. The detailed accounts of these activities were later collated into analysis forms (see Appendix 4), documenting the details of the assessment activities, how the activities were organised, what the assessment was about, and what feedback was involved for further categorisation. For this study, oral feedback in classroom conversations instead of written feedback was given attention in data generation and analysis since the study mainly focused on teacher-student interactions in classroom assessment, while teachers usually provide written feedback after class. The categorisation of the activities was guided by Hill and McNamara's (2012) dimensions of classroom-based assessment, which identify different evidence that should be noted when understanding classroom-based assessment, including what is assessed, how assessment evidence is collected, who is assessed, and by whom. Among the 32 lessons observed, 328 assessment activities were identified and categorised. The categorisation started with identifying the assessors, which included teacher, self, and peer, the responses elicited by the assessment tasks, which included oral and written responses, and the assessment activities focusing on different aspects of language learning, such as reading comprehension, grammar, and vocabulary. Codes were assigned for different activities (see Table 3-1) based on Braun and Clarke's (2006) thematic analysis (discussed in the following paragraphs). After the categorisation, detailed narratives of the teachers' classroom teaching, learning, and assessment were produced. The narratives of observed assessment moments pertinent to the research questions were included as a part of building the analytic summary of the teachers' activities and were used as data for further analysis.

For the second level of analysis, semi-structured interviews sought to elicit what teachers believed and understood about assessment activities, concepts and policies and provide data for the 'why' research questions. To analyse the transcripts of the interviews, thematic analysis was used. According to Braun and Clarke (2006, p. 79), thematic analysis is "a method for identifying, analysing and reporting patterns (themes) within data." For Braun and Clarke (2006), a theme represents something significant about the data and the research questions. By describing and organising the themes in the data, researchers can identify and interpret the patterned responses and meanings within each case and across the dataset (Braun & Clarke, 2006; Braun et al., 2019). According to Braun et al. (2019, p. 843), there are different types of thematic analysis: a "coding reliability" approach, a "codebook" approach, and a "reflexive" approach. The coding reliability approach and the codebook

approach reflect a more objectivist/positivist stance. For these approaches, themes or patterns can be determined at the start of the analytic process. Unexpected but important data is at risk of being dropped for not fitting within any pre-determined themes. On the other hand, the reflexive approach reflects a more constructionist stance. For this approach, themes or patterns are identified later in the analytic process. The researcher takes an active role in using this approach for knowledge production, and themes are constructed through understanding meanings across the dataset.

Assessor	Response	Assessment process	Focus of assessment activity	Code
Teacher	Oral response	Teachers raised planned or improvised questions, invited students to make oral responses, and gave feedback to students.	Reading comprehension	TO-RC
			Grammar	TO-G
			Vocabulary	TO-V
			Translation	TO-T
			Reading aloud	TO-RA
			Recitation	TO-R
			Speaking	TO-S
	Written response	Teachers assigned written tasks for students, asked students to finish in class or after class, and marked their written responses after class or explained the tasks in class.	Dictation	TW-D
			Write from memory	TW-WFM
			Writing exercise	TW-WE
			Reading exercise	TW-RE
			Grammar exercise	TW-GE
			Listening exercise	TW-LE
			Speaking exercise	TW-SE
			Vocabulary exercise	TW-VE
Self	Written response	Students assessed their own written responses.	Reading exercise	SW-RE
			Dictation	SW-D
			Grammar exercise	SW-GE
Peers	Oral response	Students assessed their peers' oral responses.	Reading aloud	PO-RA
			Grammar	PO-G
			Reading comprehension	PO-RC
			Speaking	PO-S
			Vocabulary	PO-V
			Translation	PO-T
	Written response	Students assessed their peers' written responses.	Reading exercise	PW-RE
			Writing exercise	PW-WE
			Grammar exercise	PW-GE
			Vocabulary exercise	PW-VE

Table 3-1 Classification scheme for classroom-based assessment activities

As the study adopts a constructionist epistemological position, a reflexive approach was taken as the analytic method for analysing interview data. The approach was conducted to first identify the within-case pattern of each case. Along with the theories and literature reviewed, the interview data were analysed individually in relation to the corresponding observation data within each case, and a narrative summary was produced for each case. Then, the interview data were compared and synthesised across cases so that a broader narrative could be framed and findings produced. Braun and Clarke (2006) and Braun et al. (2019) argue that there are six phases in reflexive thematic analysis: familiarisation, generating codes, constructing themes, revising themes, defining themes, and producing the report. The first phase is familiarisation. This phase aims to achieve a thorough understanding of the data generated. By listening to the audio data, reading textual data, and reviewing field notes, I became familiar with the depth and breadth of each case, then the whole dataset. According to the research questions, new notes were also made for preliminary inductive coding.

The second phase is generating codes. This phase aims to identify meanings and produce initial codes from the data. There are two levels at which meaning can be identified and coded: “semantic” and “latent” (Braun & Clarke, 2006, p. 88; Braun et al., 2019, p. 853). Semantic codes capture the explicit meaning of data, while latent codes focus on the implicit meaning and theoretical resonances. Initial coding is often semantic, but when analysis develops, latent meaning will emerge and be identified (Braun et al., 2019). During the initial data generation process, all data items were given full attention and coded for as many potential themes as possible to provide enough information for later interpretative analysis. The third phase is constructing themes. This phase aims to analyse the codes and consider how different codes might be classified under an umbrella theme. At the end of this phase, the umbrella themes, the sub-themes and the extracts of all the data formed a thematic map showing the relationship between codes, between themes, and between different levels of themes.

The fourth and fifth phases are revising and defining themes. These phases aim to define each theme clearly and construct the overall narrative the themes tell about the data. Reviewing data extracts was useful at this stage. The extracts for each theme were compared and synthesised to see whether they could produce a coherent pattern, and a thematic map that could reflect the entire dataset was created (see Table 3-2). The final phase is producing

the report. This phase is not purely a write-up process but a final analysis stage and a final test of how well the themes work. By framing the report, it was ensured that the final themes remained coherent with the data. The report went beyond descriptions of the data and made arguments supported by transcripts and/or translations of selected extracts to answer the research questions. For the source of the extracts, codes were provided according to teachers' codes, page numbers of the interview transcripts, or lesson number (for example, T1:10; T2:L2). For data generated during informal interviews, the extracts were coded as 'informal interview' (for example, T3:informal interview).

Theme	Sub-theme	Code
Teacher profile	Teacher	Teaching experience
		Professional education in general
	School	Class size
		Students' English language level
		Classroom layout
Teacher's understandings	Language learning & teaching	Intended teaching objectives
		Intended teaching arrangements
	Language assessment	Views of assessment activities
		Views of formative assessment
Teacher's experiences	Experiences regarding assessment	Professional learning opportunities in assessment
		Present experiences with classroom assessment
		Future expectations regarding assessment
	Experiences regarding assessment policies	Teacher evaluation at school level
		Reform in English Zhongkao test design
		Experiences with curriculum standards

Table 3-2 Thematic map

After analysing each case, cross-case synthesis will be conducted to develop plausible and strong arguments that are supported by the data, including the data obtained from both observations, interviews, and document analysis. Firstly, within-case patterns in all cases were examined whether there were relationships replicated across the cases in terms of "literal replication", which means the cases selected are similar and are possible to produce similar findings, and "theoretical replication", which means the cases hold contrasting findings (Yin, 2018, pp. 287-288). The codes and themes that helped analyse the 'what', 'how', and 'why' questions in each case were considered for the cross-case comparison. Secondly, it was helpful to review the theoretical themes relevant to the study, namely assessment activities, teachers' understandings of assessment, and assessment policy

enactment (as discussed in section 3.3.2, p. 55-56), to interpret the meaning of the emerging literal and theoretical relationships. For literal relationships, it was necessary to discuss how individual cases were comparable according to different dimensions to elicit commonalities between them. Similarly, for theoretical relationships, the marked differences among cases added another layer to the complexity of the research outcomes.

The validity and reliability of this study were also considered throughout the data analysis process. Creswell and Miller (2000, pp. 124-125) argue that validity in qualitative research is defined as "how accurately the account represents participants' realities of the social phenomena." Gibbs (2007, p. 97) suggests that reliability in qualitative research indicates that the approach is "consistent across different researchers and different projects." Several strategies could be incorporated into the study to address validity and reliability in the research process. According to Kvale (1989, p. 77), "validation becomes investigation, continually checking, questioning, and theoretically interpreting the findings." Here, validation is not a final verification. It is "built into the research process with continual checks of the credibility, plausibility, and trustworthiness" of the research outcome (Kvale, 1994, p. 168). In both within-case analysis and cross-case analysis of this study, the data were continuously reviewed, compared, and synthesised to ensure that the themes constructed were coherent and the arguments made were strong. After establishing the preliminary themes, I also searched through the documentary, observational, and interview data for evidence that disconfirmed these themes. Such a procedure could serve as a triangulation method to strengthen the trustworthiness and credibility of the analysis, as it can be used to make sure that the themes constructed truly echo the collected data and reflect the complex realities observed (Creswell & Creswell, 2018; Creswell & Miller, 2000).

Rowe (2014) also argues that the positionality of the researchers could bring influence to their study. Positionality refers to "the stance or positioning of the researcher in relation to the social and political context of the study" (Rowe, 2014, p. 628). It is closely connected to the researcher's background and is likely to affect every aspect of the study. For an interpretivist enquiry, the researcher should address their positionality and reflect on the research process about how the outcomes of the study may be affected and how an open and honest conversation with the reader can be created. For this study, data analysis is conducted through an insider's eyes: my experiences as a student in the school systems in Shenzhen, a language teacher who worked in Chinese institutions, and a curriculum designer who

designed language learning materials for Chinese students are drawn on to better understand the teachers' narratives, their working environment, and their positions in their working environment.

Besides letting researchers have control of the validity check, Gibbs (2007) suggests that the participants should also be involved in determining the accuracy of the outcomes. According to Gibbs (2007), researchers can present a piece of transcript, the derived themes, the major outcomes, and so forth, to the participants, and provide them with the opportunities to confirm, comment, or change what they have said during the data generation process. This can serve as a follow-up interview to make sure that the participants agree with the transcripts and the interpretations. Creswell and Creswell (2018) also argued that providing a chance for the participants to comment on the final report, specific descriptions, or themes of the study could be a crucial technique to establish credibility and make sure that the findings are accurate. For this study, moments of assessment observed in the observations and transcripts of interviews were shared with the participants in the after-observation interviews with the opportunities to comment or to change any statement, and the narratives of each case were provided to the participants so that they could confirm the narrative accounts.

As for reliability, several procedures are also suggested by researchers to ensure the consistency and quality of the research approaches. Gibbs (2007) argues that transcripts should be checked thoroughly so that no obvious mistakes are involved. He also suggests that the process of coding and constructing themes should be paid special attention to so there is no shift in the definition of codes and themes across the dataset. Constant comparisons within a case and between cases were therefore conducted in this study. Yin (2014) makes specific suggestions for researchers conducting a case study. He argues that researchers should keep records of as many of the steps of the procedures in their case studies as possible for the convenience of future analysis and inspection. This study followed the suggestions and kept the documents of classroom observation schemes, interviews transcripts, and analysis forms for the observed assessment activities to ensure that the data could be reviewed back and forth when needed.

3.4 Ethical considerations

This study follows the British Educational Research Association (BERA) (2018) guidelines for ethical research in education and obtained ethics approval (see Appendix 5) from the College Research Ethics Committee of the University of Glasgow. In this section, I explain how participants in my study were protected and how the ethical conduct of research was guaranteed in my research.

According to Goodwin et al. (2003, p. 567), "ethics is an ever-present concern for all researchers." Ethics pervades the entire research process, from design to practice, and then to analysis, writing, and dissemination. It is a pressing concern for all qualitative researchers due to the "emergent, dynamic and interactional nature of most qualitative research" (Iphofen & Tolich, 2018, p. 1). Educational research is no exception. Over recent decades, increasing focus has been put on ethical issues within educational studies. Legislative changes have taken place in many countries to regulate how data are managed and stored; many research funders have established new ethical rules; new technologies have introduced new research methods, which give rise to new ethical challenges (Brooks et al., 2014). The legislative and regulatory requirements have made conducting ethical research imperative for all educational researchers. It should not be ignored that conducting ethical research is also important in itself. As Iphofen and Tolich (2018) argue, the values that lie behind the principles and standards of research inform how we conduct research. The values we hold, such as respect, equality, and dignity, will influence our behaviours and attitudes in conducting research and will eventually influence the conclusions we reach.

Educational research aims to achieve knowledge and understanding regarding all modes of educational activities from all stakeholder perspectives, including learners, teachers, policymakers, and the public. The research process usually involves human participants, and the data obtained in the research is usually related to the lives and activities of these people. If the research activities are not carefully planned and examined, the participants could be harmed psychologically or even physically during the process; if the research data is not well managed and stored after generation, the rights, privacy, and dignity of the participants may also be potentially harmed by data leakage or misuse of their information. As a result, educational researchers are responsible for carefully planning their approaches to data generation and storage to avoid any harm to the participants and the public.

Since the study involves interviewing teachers and observing teachers' classrooms, the main participants in the study are identified as teachers. During the research process, the teachers were treated "fairly, sensitively, and with dignity and freedom from prejudice" (BERA, 2018). Before the data generation, to ensure that the teachers were fully aware of the nature of the study and the processes they engaged in, the teachers were provided a teacher participant information sheet (see Appendix 6) to inform them about the aim and details of the study and a consent form (see Appendix 7) to obtain their permissions to conduct classroom observations in their classrooms and interview them. Consents were also obtained from the schools where the research was undertaken (see Appendix 8) and from the students in the observed classrooms along with their parents (see Appendix 9). The right of participants to withdraw from the study at any time and for any reason was also preserved. If they felt uncomfortable about the questions in the interviews or were unable to involve me as an observer in their classrooms, they could inform me about the situation, and I would cease asking the questions or observing and consider whether a change of approach was needed.

The research process, especially the interviews, was carefully designed to avoid potential risks and minimise and manage any distress, inconvenience, or discomfort for teachers. The study involved observing the teachers' teaching and assessment and investigating their educational values and beliefs. The presence of a researcher might cause uncomfortable feelings of being watched or judged during observations and interviews. To tackle this issue, I encouraged the teachers to treat me as a learner, not a supervisor or inspector, so they understood there were no repercussions for them. In addition, interview transcripts and moments of assessment were shared with the participants with the opportunity to change their comments. The time and effort needed for participating in the study were also considered, as teachers are usually busy with their teaching, lesson preparation, marking, and other administrative work. In order to minimise the impact of my research on their workload, I negotiated with teachers about their timetables before conducting my research.

Privacy and data storage are also important parts of my ethical considerations. Data collected in this research, including audio recordings and transcripts, are used for this project only. The identities of the schools and teachers were anonymised throughout the research and will not be revealed on any other occasion. The confidentiality of the teachers' personal information, such as name, age, and workplace, was ensured by excluding any identifiable

details in the interview or observational data. I also complied with the legal requirements regarding the storage and use of the data obtained as specified in the General Data Protection Regulation (GDPR) (Council of the European Union and European Parliament, 2016). The participants are entitled to know how, where, and why their data is being stored and who will have access to it.

Chapter 4 Materials analysis

In this chapter, the textbook and teacher guidebook (one textbook set in a textbook series, introduced in section 4.1) used by the participants and the English Zhongkao test design (section 4.2) will be analysed with a content analysis approach (Bowen, 2009). This analysis explores the assessment ideas underlying the textbook set and the Zhongkao with an inductive analysis approach, which provides the background context and a source of reference for the analysis of the two cases. The official curriculum standards introduced in Chapter 1 will be addressed to characterise the connection between the policy, textbook, and the English Zhongkao. Concerns regarding teachers' assessment practices, textbook, and English Zhongkao will be proposed for further exploration in classroom observations and interviews.

4.1 Textbook and teacher guidebook

As the representation of the official curriculum, the textbook and teacher guidebook used by the participants of this study need special attention. The textbook series used in middle schools in Shenzhen is mandated by the local educational authority. Co-published by the Shanghai Educational Publishing House and the Oxford University Press, the textbook is designed and edited according to the 2011 'English Curriculum Standards for Compulsory Education' (Ministry of Education, 2011). The series consists of six textbooks: English for Grade 7 (volume 1), Grade 7 (volume 2), Grade 8 (volume 1), Grade 8 (volume 2), Grade 9 (volume 1), and Grade 9 (volume 2). Each textbook is accompanied by a teacher guidebook, which involves detailed guidance for teachers, introducing the textbook structure, teaching suggestions for each unit, recommended classroom activities, tape scripts, answer keys for the exercises, and suggested teaching plans.

For each textbook, there are four modules, and each module consists of two units. Each module represents a broad topic, and each unit represents a specific sub-topic related to the broader one. For example, the second module in English for Grade 8 (volume 1), which was the textbook used by the participants during the fieldwork, is called 'Science and technology'. This module has two units, which are called 'Computers' and 'Inventions'. For each unit, there are ten sections: 'getting ready' (section 4.1.1), reading (section 4.1.2), listening (section 4.1.3), grammar (section 4.1.4), speaking (section 4.1.5), writing (section 4.1.6), 'more practice' (section 4.1.7), project (section 4.1.8), 'cultural corner' (section 4.1.9), and

self-assessment (section 4.1.1). Apart from ‘more practice’ and ‘project’, which are optional, all other sections are routine sections, which teachers should address in every unit. The design of each section will be illustrated later in this chapter.

The guidebook suggests that teachers should take approximately eight lessons to teach one unit over two weeks. For each lesson, a suggested detailed teaching plan is provided in the guidebook. However, this is only for reference – teachers need to “adjust their teaching design according to their teaching context” (Zhang & Shu, 2013, p. vi). The number of lessons assigned for each section in a unit is also suggested by the guidebook. This suggestion does not reflect how teachers actually plan their teaching, but it may carry implications of which section is considered more important by the textbook designers. A list of suggested time for one unit and different sections is displayed in Table 4-1:

Section	Lessons	Lessons for one unit
Reading	2.5	8
Listening	0.5	
Grammar	1.5	
Speaking	0.5	
Writing	0.5	
More practice	1	
Project	1	
Cultural corner and self-assessment	0.5	

Table 4-1 Suggested time for one unit and different sections (Zhang & Shu, 2013, p. vi)

Amongst all the sections, ‘reading’ and ‘grammar’ are allocated more time (2.5 and 1.5 lessons respectively). ‘More practice’ and ‘project’ are allocated one lesson, while ‘listening’, ‘speaking’, ‘writing’, and ‘cultural corner and self-assessment’ are each allocated 0.5 lessons. It should be noted that the reason why ‘reading’ is given excessive time may be that vocabulary teaching is usually integrated with teaching the reading passages. Still, it can be inferred that ‘reading’ and ‘grammar’ are considered the more important elements in each unit.

Although the textbook is organised by topics, it is not designed as entirely communicative oriented; rather, it seeks to strike a balance between teaching language knowledge and teaching communication. Throughout the guidebook, the textbook designers emphasised consistently that students should acquire the necessary language knowledge and use the knowledge to accomplish certain tasks. For instance:

> Grammar is an important part of middle school students' English learning. All grammar points are presented in the reading passage, and are practised and consolidated in listening, speaking and writing activities. (in 'grammar' section, Zhang & Shu, 2013, p. iii)

> Students should use the vocabulary and grammar knowledge in the unit to hold discussion and express their opinions based on the topics. (in 'speaking' section, Zhang & Shu, 2013, p. iii)

The guidebook seems to follow an "eclectic approach" (Brown, 2000, p. 40), which gives general attention to both form-focused pedagogy and meaning-focused pedagogy, two streams of approaches that can be considered stemming from behaviourist and communicative teaching perspectives respectively. This corresponds with the curriculum's philosophy of design, which stresses the significance of both language meanings and forms (see section 1.3, p. 9 for detailed description).

In the following sections, I will describe and analyse the design of each section of the textbook unit in 'English' for Grade 8 (volume 1), which is the textbook used by the participants at the time of data generation. Specific attention will be given to unit 2 and 4 (see Appendix 10 for the two textbook units), as the teaching of these units were observed in the two schools. The explanation and suggestions made in the guidebook for each section will also be addressed. Special attention will be given to the assessment practice recommended by the guidebook. The opportunities made available for summative and formative assessment in the textbook design will also be attended to in terms of how they reflect formative assessment principles.

4.1.1 Getting ready and self-assessment

Each unit of the textbook starts with a 'getting ready' section (see Appendix 10, p. 246/262) and ends with a 'self-assessment' section (see Appendix 10, p. 259/275). The two sections outline a list of learning objectives of the unit for different purposes. The guidebook

emphasises that ‘getting ready’ “helps students understand the learning objectives and corresponds with the ‘self-assessment section’ at the back” (Zhang & Shu, 2013, p. i), while ‘self-assessment’ allows students to “self-assess themselves at the end of the unit, find the gap between their performance and the objectives and improve their learning” (Zhang & Shu, 2013, p. iv). In the self-assessment list at the end of each unit, three facial expressions (smiley face, neutral face, sad face) were given, following each objective. Students are expected to tick the boxes to indicate how well they feel they have grasped the knowledge of this unit.

The guidebook frames the two sections as an assessment pack that helps students to check on their own progress. From a formative assessment point of view, the two sections manage to clarify and share learning objectives and criteria for success (Black & Wiliam, 2009). They help establish where the students are in their learning, where they are heading, and what needs to be done to accomplish their goals (Wiliam & Thompson, 2007). However, there is no further pedagogical guidance in either textbook or guidebook on what teachers and students should do if students were to tick neutral faces or sad faces in the self-assessment list. Neither are teachers prompted by the guidebook to check how students respond to the objectives or to provide students with further feedback and support with the faces they ticked. Whether and how teachers and students make use of these two sections will be further investigated in classroom observations and interviews.

‘Getting ready’ also has another small routine section, which is a cartoon related to the unit’s topic. The conversation of the cartoon characters involves language points related to the topic of the unit or the different meanings or usages of a particular word. The guidebook suggests that the cartoons could “help teachers introduce the topic of the unit” and “stimulate students’ curiosity about the topic” (Zhang & Shu, 2013, p. i). Teachers are also advised to raise questions about the cartoon or ask students to share their experiences related to the conversation. This provides teachers with opportunities to elicit students’ understanding of the cartoon and could assist their further instructions so that students could better understand the topic of the unit.

4.1.2 Reading

The second routine section is ‘reading’ (see Appendix 10, p. 247/263). This section consists of five components: a lead-in component called ‘what do you know about …?’, a pre-reading component called ‘before you read’, a reading passage, and two exercise components,

'vocabulary' and 'comprehension'. The first component, 'what do you know about …?' (see Appendix 10, p. 247/263), is a preparatory task(s) that involves matching, discussion, gap-filling, or multiple-choice. As instructed by the guidebook, this component includes "warm-up exercises" that "help students understand the topic of the unit" (Zhang & Shu, 2013, p. i). The task(s) aim to elicit "students' prior knowledge and their own life experience connected with the topic" and help them "clear out some of the difficulties that impede their understanding" (Zhang & Shu, 2013, p. i). According to Jones and Saville (2016, p. 41), if a classroom activity requires certain prior knowledge, "a first step is to begin with a check on whether students possess that knowledge." It would be interesting to observe whether and how teachers address what students already know in classroom conversations.

The guidebook suggests that, for the 'what do you know about …?' component, teachers can lead an oral discussion about the task(s) or organise group discussions among students. This pedagogical design, as argued by the guidebook, can "arouse students' interest in the reading passage and the topic" and, therefore, "arouse students' interest in learning English" (Zhang & Shu, 2013, p. i). After eliciting students' prior knowledge of the topic, teachers can teach them new words, introduce other background knowledge, or ask students to research themselves on the topic. While the task(s) and the suggested pedagogy could be used to assess students' prior knowledge and draw evidence for future teaching planning, the guidebook does not mention that teachers can use the task(s) for assessment purposes. The statement that has the most assessment implication is "(the section aims to) clear out the gaps that might impede students from understanding the text" (Zhang & Shu, 2013, p. i), but there is no further explanation of how teachers can achieve such an aim through formative interactions. It seems that the guidebook describes this section as a section that promotes students' motivation, which has more pedagogical function rather than assessment one. While explaining the rationale of the activities, the guidebook argues that the activities could "activate (激活) students' background knowledge" (for example, Zhang & Shu, 2013, p. i). The premise seems to be that the background knowledge of students is already there, and all students share the same knowledge base. All teachers need to do is address this knowledge and move on to the next stage.

The second component, 'before you read' (see Appendix 10, p. 247/263), is a pre-reading component that encourages students to practise skimming reading skills. This component typically asks students to predict the content of the passage by looking at the pictures, titles,

sub-headings, and the first sentence of every paragraph of the passage. The task(s) in this component could be matching, answering questions, completing tables, and ticking the right answers. The guidebook argues that this section could "increase students' learning desire, improve their comprehensive ability and prepare themselves for the following sections" (Zhang & Shu, 2013, p. i). Teachers are advised to give more initiative to students and encourage them to "actively find the answers in the passage", "even though mistakes could be made in the process" (Zhang & Shu, 2013, p. ii). According to the guidebook, teachers are suggested to "encourage students to check their answers for the task(s) themselves while reading the passage" (Zhang & Shu, 2013, p. ii). Opportunities for errors are provided "as errors are necessary for learning" (Jones & Saville, 2016, p. 42), and opportunities for students' self-assessment are also opened up.

The guidebook further suggests that teachers could "ask students to make a comparison between the answers they predict in pre-reading and the answers they find while reading" (Zhang & Shu, 2013, pp. i-ii). This could potentially encourage students to reflect on their reading strategies and allow teachers to assess students formatively by noting the commonly made mistakes and referring to them while teaching the reading passage. However, the guidebook does not mention that teachers could use the information drawn from students' comparison of answers. It only stresses that this could encourage students to learn more actively.

The third component is the reading passage (see Appendix 10, p. 248/264), the core component of the whole reading section. The reading passages selected for Grade 8 usually involve 180 to 200 words. They incorporate different writing genres: some are presented as stories, some as reports, and some as posts on the Internet. The passage involves the new vocabulary and grammar points outlined in the learning objectives, allowing students to learn such language knowledge in the text. The guidebook provides teachers with a background story, the main teaching points, sample activities, and a list of new words and expressions. Teachers are encouraged to link back to the reading strategies practised in 'before you read' while teaching, which hypothetically creates opportunities for teachers to provide feedback to students regarding their performance in the last component. Teachers are also suggested to pay additional attention to the language knowledge in the text, which is addressed by the exercises in the following two components, 'vocabulary' and 'comprehension'.

The fourth and fifth components, ‘vocabulary’ and ‘comprehension’ (see Appendix 10, pp. 249-250/265-266), take the form of exercises such as multiple-choice, gap-filling, and short-answer questions. According to the guidebook, these two components aim at sharpening students’ understanding of the reading passage: ‘vocabulary’ “helps to foster students’ ability to use the new words and expressions,” while ‘comprehension’ “leads students to a more in-depth understanding of the reading passage” (Zhang & Shu, 2013, p. ii). The guidebook suggests that “it is better for students to finish the exercises in ‘vocabulary’ in pairs or groups, and teachers should check students’ answers in class” (Zhang & Shu, 2013, p. ii). This could allow students to collaborate with their peers and let teachers know how well students have understood the words. The guidebook also suggests that exercises in ‘comprehension’ can be finished “individually, in pairs or in groups” (Zhang & Shu, 2013, p. ii). It seems to recognise the benefits of student cooperation and advocate peer dialogues around learning, which can activate different sources of feedback and strengthen students’ abilities to become self-regulated learners (Nicol & Macfarlane-Dick, 2006).

The exercises in ‘vocabulary’ and ‘comprehension’ adopt “a layered design” (Zhang & Shu, 2013, p. ii), which begins with easier tasks requiring simple answers, and follows with more difficult tasks requiring more complicated answers. According to the guidebook, this design “allows students to understand the text step by step” (Zhang & Shu, 2013, p. ii). From an assessment perspective, arranging the task with different difficulty levels allows the construction of a task that enables effective scaffolding (Jones & Saville, 2016; Rogoff, 1997). Students may review the reading passage along with the vocabulary while completing the tasks and rethink the deeper meaning of the text through the given tasks. This design could also enable a formative use of summative assessment, as it can communicate to students what knowledge is valued and elicit students’ achievement (Black & Wiliam, 2009). If used appropriately, teachers could provide feedback based on students’ performance on these tasks to move their learning forward. Whether and how teachers provide feedback should be explored in classroom observations.

To summarise, there are opportunities for summative and formative assessment practice built into the entire ‘reading’ sections. More progress- and performance-based assessment, which is encouraged by the official curriculum, has been evidenced in the textbook and teacher guidebook. However, there is the empirical question of how these resources are used and enacted in actual classrooms, which awaits to be discovered in classroom observations and

interviews. It should be noted that most of the assessment tasks presented in this section adopt the form of discrete-point test, which resembles the formal examinations students take. On the one hand, it is reasonable for the textbook to adopt such forms, as it provides an opportunity for students to practice their test-taking skills. However, on the other hand, it could also make it easier for students and teachers to ignore the formative use of these tasks.

4.1.3 Listening

The third routine section is 'listening' (see Appendix 10, p. 251/267). According to the guidebook, this section offers students a listening task(s) to learn and practise listening skills. The tasks share the same topic with the reading passage and incorporate different types of recordings, such as conversations, broadcasts, interviews, and stories. The grammar points outlined in the objectives and learned in the reading passage are presented in the tasks and, as the guidebook argues, "serve as the groundwork for the next section 'grammar'" (Zhang & Shu, 2013, p. ii). This enables the recurrence of grammar points, allowing more feedback and scaffolding opportunities (Rea-Dickins, 2006). The guidebook suggests that teachers should "clearly inform students of the requirements of the listening tasks" so that students are informed of the goals and criteria of this section (Zhang & Shu, 2013, p. ii). It also emphasises that teachers should "guide the students to use the appropriate skills to complete the tasks", such as skimming the text and predicting through text and pictures (Zhang & Shu, 2013, p. ii). These suggestions emphasise sharing criteria with students, which resembles one of the formative assessment strategies proposed by Black and Wiliam (2009). They could inform students about what counts as good performance and help them develop their self-assessing skills (Jones & Saville, 2016).

From the analysis so far, it is evident that grammar is considered an important learning objective since it is addressed in both 'reading' and 'listening', not to mention there is another 'grammar' section in the unit. It is reasonable for the textbook design to stress grammar learning, as grammar is an important part of language learning, and the lack of knowledge in grammar could restrict meaningful communication (Canale & Swain, 1980). However, as discussed in section 2.1.2, there are both pros and cons in grammar-focused teaching. A primary con of over-emphasising grammar in language teaching is leading to "insistence on learner accuracy" and "a frustration for teaching to communication" (Savignon, 1991, p. 267). This might contradict the objective of the curriculum, which aims to develop students' practical language abilities (Wang, 2007). There are also opportunities

for both summative and formative assessment, but again, from the guidebook, there is little trace of encouraging formative assessment in this section. The suggestions made by the guidebook seem more relevant to pedagogical procedures and do not address how teachers can elicit information from students' performance and provide further guidance.

4.1.4 Grammar

The fourth routine section is 'grammar' (see Appendix 10, pp. 252-254/268-270). This section consists of explanations for target grammar points and exercises for each grammar point. According to the guidebook, the grammar points in this section "have already appeared in the reading passage and the 'listening' section" and will also be "practised in the following two sections: 'speaking' and 'writing'" (Zhang & Shu, 2013, p. iii). The exercises in this section are related to the topic of the unit so that students can have easier access to the understanding of the tasks. The forms of these exercises and the activities recommended by the guidebook are mostly discrete-point tests such as gap-filling and making sentences, which focus on the practice of the target grammar points. Other skills, such as speaking and writing, are sometimes practised in this section but comprise a small proportion of the tasks. The pedagogical guidance for this section mainly consists of descriptions and examples of the grammar rules. Some practical advice is assigned accordingly to different units, such as using teaching aids to demonstrate the grammar rules or asking students to role-play.

Besides a few lines in some of the units suggesting that teachers can comment on students' performance on the suggested exercises and activities, the guidebook does not address how teachers should assess students formatively while teaching grammar and using the given exercises and activities. There could be opportunities for formative assessment though. Since the target grammar points are practised across all main sections, teachers should have ample opportunities to check on students' grammar learning and provide timely feedback for correction and improvement. The repeating practice of grammar points could help students better memorise the grammar rules and master the grammar points more comprehensively through learning all the sections (Brown & Lee, 2015; Jones & Saville, 2016). The guidebook, however, seems to give teachers the initiative to use the given tasks and assess students in grammar. Classroom observations should further explore what assessment practices teachers adopt in this section.

4.1.5 Speaking

The fifth routine section is ‘speaking’. This section consists of two components: ‘talk time’ and ‘speak up’. The first component, ‘talk time’ (see Appendix 10, p. 255/271), focuses on practising pronunciations, including phonetic alphabets, intonations, and weak and strong forms. Rules of pronunciation are introduced, and pronunciation exercises are displayed. The guidebook suggests that, in this component, teachers should “focus on correcting students’ pronunciation and intonation” and “teach them commonly used expression with the functional-notional approach (功能意念教学)” (Zhang & Shu, 2013, p. iii). According to Brown and Lee (2015), the functional-notional approach focuses on the pragmatic purposes of using a language. The ‘functional’ part corresponds to the actual use of language for meaningful communication, while the ‘notional’ part focuses on the context and situation where the conversations occur. This proposed approach matches what is claimed by the curriculum objectives, as it embraces ‘authentic’ communication practices. However, the ‘talk time’ section in which the approach is proposed does not include any conversational practice in a given situation. It only requires students to read the given words and sentences correctly but does not provide opportunities for them to practice speaking under a certain topic. The connection between the textbook and the suggested pedagogy seems weak, and teachers can only assess pronunciation instead of speaking skills if they follow the given tasks.

The design of the second component, ‘speak up’ (see Appendix 10, p. 256/272), however, seems to reflect the functional-notional approach, even though the guidebook does not explicitly suggest the approach for this component. It is uncertain whether this is a typo. The ‘speak up’ section aims to encourage students to “use the knowledge they learned in this unit”, such as vocabulary and grammar, to “hold conversations among their peers” and “express their opinions” (Zhang & Shu, 2013, p. iii). For this component, a task is assigned to students to perform in pairs or groups. Examples and hints are given as scaffolding. According to the guidebook, teachers should “encourage every student to participate in the conversation”, and students should “make full use of the vocabulary and sentence structures they learned” in the task (Zhang & Shu, 2013, p. iii).

The design of the second component allows more opportunities for teachers to assess students formatively. The guidebook makes various suggestions for this component, asking teachers to motivate students to speak up. For example, the guidebook suggests that teachers

could “guide students to finish the conversation tasks after explaining the grammar section so that students can learn how to use the grammar knowledge they learnt” (Zhang & Shu, 2013, p. 39). This addresses students’ previous learning process and outcome, allowing for ongoing learning, assessment, and instruction (Shepard, 2000). The tasks in ‘speak up’ also require students to collaborate. This allows students to engage in self- and peer assessment. It should be noted that although the curriculum claims that it aims to improve students’ communicative competence, the guidebook suggests that teachers should spend 0.5 lessons on the speaking section. It is odd that the guidebook only allocates such little time for the speaking section, and it would be necessary to check how much time the participants actually spend in the speaking section in classroom observations.

4.1.6 Writing

The sixth routine section is ‘writing’ (see Appendix 10, p. 257/273). This section aims at “training (训练) students’ writing skills” and “learning how to write in different styles, for example, postcard, story, note, and articles” (Zhang & Shu, 2013, p. iii). This section usually involves different tasks with different difficulty levels, and each task serves as scaffolding for students to imitate and learn how to write properly. For example, in unit 2, Grade 8 (volume 1), the ‘writing’ section asks students to write a report based on a line graph. The section is divided into two tasks. The first task shows a line graph and asks students to complete a cloze with numbers and months according to the figures in the graph. This helps students familiarise themselves with “the key expressions and sample writing”. The second task is to write their own report. Another line graph is presented, and students are required to write the report following the text in the first task. The construction of the two tasks shows clear consistency, which provides students with proper scaffolding to move forward (Jones & Saville, 2016). Students can learn what a line graph is and how to write a report based on the graph with the teacher’s guidance and the examples given in the textbook.

For this section, the guidebook suggests that, while teaching this section, teachers should “cultivate students’ ability to use the language knowledge they learned previously” (Zhang & Shu, 2013, p. iii). It also suggests that students should “revise their classmates’ reports in pairs” (Zhang & Shu, 2013, p. 40), creating opportunities for students to peer assess each other’s work. This suggestion intends to have students engage more in peer assessment, which could help students improve their own writing and promote their understanding of

writing criteria and learning goals. Whether and how teachers follow these suggestions should be observed in the fieldwork.

4.1.7 More practice

The seventh section, ‘more practice’ (see Appendix 10, p. 258/274), is an optional section in each unit. Teachers can choose to teach this section if there is more time in class or assign it as homework for students to finish after class. This section consists of an additional reading passage related to the unit’s topic, reading comprehension exercises, and a discussion task. Again, the unit’s vocabulary and grammar points are articulated in the passage. Unlike the passage in the ‘reading’ section, this passage is for extensive reading rather than intensive reading, so the guidebook suggests that teachers do not need to teach the passage word by word, sentence by sentence. Instead, “understanding the story actively” matters more (Zhang & Shu, 2013, p. iv).

There is no mention of assessment in the guidebook about this section, although this section seems like a reading exercise that needs marking. However, there could be opportunities in this section for teachers to assess students formatively if students are required to answer the questions or express their opinions during a discussion in class. The interesting thing is, although the guidebook suggests that ‘more practice’ is an optional section, according to the suggested time for different sections (Zhang & Shu, 2013, p. vi), this section is allocated one lesson, which is longer than the time assigned for speaking and writing sections. It seems that reading indeed receives more emphasis from the textbook designers.

4.1.8 Project

The eighth section is ‘project’ (see Appendix 10, pp. 260-261/276-277), which is an optional section that encourages “inquiry-based learning” (探究性学习) (Zhang & Shu, 2013, p. iv). For this section, a project related to the topic of the module is assigned to students, asking them to collaborate with their peers, and finish the project by “discussing, investigating, and researching for information from books, newspapers, media, and the Internet” (Zhang & Shu, 2013, p. iv). For example, the ‘project’ for the first module of Grade 8 (volume 1) is ‘A mini-encyclopaedia of numbers’. The project is divided into five tasks. Firstly, the students are asked to brainstorm topics about numbers in groups and write down the topics on a given mind map. Secondly, each group is asked to decide on a topic and do some research on it.

Thirdly, each group should write an article on the decided topic following a writing sample. Fourthly, students are required to put all the articles together, make a mini-encyclopaedia, and create a table of content with the articles displayed. Fifthly, they need to design a cover for the encyclopaedia.

Throughout the tasks, guidance, mind map, hints, and sample writing are provided for students to assist their discussion and collaboration. These task elements could encourage students to participate jointly in problem-solving and focus their attention on the tasks, which could serve as effective scaffolding that supports their completion of the tasks (Jones & Saville, 2016; Rogoff, 1997). For this section, the guidebook does not make explicit suggestions for assessment practice. It only suggests that teachers should "encourage students to use their imagination and creativity" and "foster students' ability to discover and solve problems" (Zhang & Shu, 2013, p. iv).

4.1.9 Culture corner

For each unit, there is this small section called 'culture corner' (see Appendix 10, p. 259/275) presenting a piece of cultural knowledge about China, English-speaking countries, or the world. This section intends to introduce students to some of the cultural phenomena related to the topic and give students a taste of how to introduce cultural phenomena in English. For this section, the guidebook suggests some questions and activities for students to finish after reading the 'cultural corner' section. Discussion topics are also assigned for some units, and students are encouraged to discuss the topics among their peers and later present their discussion in class if, according to the guidebook, they have a high level of English. There is no suggestion for assessment in this section; teachers appear to have been left to decide whether they conduct assessment activities during questions and discussions.

So far, the textbook and teacher guidebook have been analysed in terms of their layout and underlying assessment ideas. The textbook is designed to present language knowledge and skills in a separate way, with some sections aiming to foster students' ability to use the knowledge and skills integratively. It seeks to strike a balance between form-focused teaching and communicative teaching by incorporating both form-focused exercises and tasks that require students to discuss and share their ideas. It provides opportunities for teachers to assess students formatively, but the teacher guidebook offers little advice and examples on how teachers should conduct formative classroom assessment. In the next

section, the design of English Zhongkao, the high-stake test in this context, will be analysed to explore its possible influence on teachers' work. The assessment ideas underlying the test and whether an alignment exists between the textbook and the Zhongkao will be investigated.

4.2 English Zhongkao design

The study was conducted when the English Zhongkao in Shenzhen was about to experience changes. Since the new English Zhongkao had more impact on the teacher participants and their students, this analysis will focus on the general information and the design of the new test.

According to the Zhongkao guidelines (Shenzhen Admission and Examination Office & Shenzhen Institute of Education Sciences, 2021, p. 65), the English Zhongkao is:

> a summative test at the end of the compulsory education period, which aims to reflect the level middle school graduates reach in English learning ... the result of the test is both the main reference for evaluating whether the students have reached the graduation standards and the basis for high school recruitment.

From the description, Zhongkao is used as an achievement test that judges whether students can obtain a middle school qualification and a selection test that decides whether a place in high schools could be offered to students. It claims to be designed based on the 2011 curriculum standards (Shenzhen Admission and Examination Office & Shenzhen Institute of Education Sciences, 2021), the same as the textbook used by the teachers. However, no evidence in the Zhongkao guidelines suggests that the Zhongkao is designed based on the textbook content.

The English Zhongkao in Shenzhen is worth 100 points in total and consists of two parts, a listening-and-speaking test worth 25 points and a written test worth 75 points. The balance of points will be further discussed in section 4.3 and the discussion chapter. The listening-and-speaking test is a 20-minute computer-based test that students need to take through an online automated scoring system. It is normally administered six weeks before the written test. The student candidates in the city are grouped into different cohorts and take the test at different periods within two days. Each student is assigned and works on a computer with the automated scoring system installed during the test. For the cohorts, different tests with the same test format are formulated with tasks selected from a question bank. Since the new official listening-and-speaking test conducted in recent years has not been made public, the

analysis of this test (section 4.2.1) will be based on the mock tests teachers use, which share the format with the official ones. A mock test, along with its scripts and answer keys provided by one of the teacher participants, is presented in Appendix 11.

The written test is a 70-minute paper-and-pen test. The student candidates take the test simultaneously in classrooms under the supervision of two test monitors with no access to other materials. The multiple-choice items in the test are automatedly scored, and the gap-filling and writing tasks will be double-blindly scored by teachers recruited from different schools. The written test comprises different tasks: two cloze tasks, reading comprehension tasks, and a writing task. The analysis of the written test (section 4.2.2) will be based on the official English Zhongkao written test for 2021, which was the exam paper for the students under observation in this study (see Appendix 12).

4.2.1 Listening-and-speaking test

The listening-and-speaking test includes three parts: 'imitate and read', 'extract information', and 'retell a story and ask questions'. The first part, 'imitate and read' (see Appendix 11, p. 278), requires students to listen to the recording of a piece of text (around 60 words) and read the text aloud, imitating the recording. The students are given 50 seconds to prepare after the recording and another 50 seconds to read the text. This task is worth four points and aims to examine whether students can correctly pronounce the words in the text and read from the text fluently with proper stress and intonation. Students are not expected to make up their own sentences in this task. All they need to do is read from the text. The aims of the task generally match the language level descriptors of phonetics for middle school graduates, which are enlisted in the curriculum standards and emphasise students' mastery of "pronunciation", "stress", and "intonation" (Ministry of Education, 2011, p. 19).

The second part, 'extract information' (see Appendix 11, pp. 278-279), involves two tasks. The first task is called 'listen and choose'. The students are required to listen to three short conversations and answer two questions after each conversation. They are given ten seconds to read the questions before each conversation. After listening to the conversation twice, they are given eight seconds to answer each question according to the three choices provided next to the question. The task is worth six points and seems a multiple-choice listening task at first glance, with the only difference being that students need to say the answer out loud in

a sentence instead of writing the answer down on a piece of paper. This task appears to address the level descriptors of both listening and speaking enlisted in the curriculum, which state that students should be able to "understand conversations about familiar topics and extract information and opinions from the conversation" and to "provide information based on simple topics" (Ministry of Education, 2011, p. 17).

The second task of 'extract information' is called 'answer questions'. The students are required to listen to a short passage (around 70 words) twice and answer four questions without any hints. They are given eight seconds to answer each question. This task is worth four points and seems more like a speaking task than 'listen and choose' because no choices are provided. Students need to take notes on the information they heard and respond to the questions accordingly. The task focuses on students' skills "to take simple notes on what they heard" and "communicate information", which are included in the level descriptors in the curriculum (Ministry of Education, 2011, p. 17).

The third part, 'retell a story and ask questions' (see Appendix 11, p. 279), also includes two tasks. The first task is called 'retell a story'. The students are required to listen to a story (around 80 to 90 words) twice and retell it in 60 seconds. A mind map is provided, and students are given 15 seconds to familiarise themselves with the hints offered by the mind map and 60 seconds to prepare for the retelling after the recording. A sentence to begin with is provided. This task is worth eight points and aims to examine whether students can "understand stories told at a natural speed" (Ministry of Education, 2011, p. 17) and "tell simple stories" (Ministry of Education, 2011, p. 15). It seems much more difficult than the previous tasks, considering the length of the recording and how much information students should reproduce. It is not surprising to see how much scaffolding this task supplies to assist students' performance.

The second task is called 'ask questions', which is based on the topic of the story in the first task. The students are required to ask the storyteller two questions. The contexts of the questions are illustrated in Chinese, which might serve the purpose of better conveying the meaning of the test instructions. For each question, students are given 15 seconds to prepare and eight seconds to respond. This task is worth three points and examines whether students can "ask for information effectively" (Ministry of Education, 2011, p. 17).

All in all, the listening-and-speaking test attempts to assess students' listening and speaking skills with a range of different tasks. The 'imitate and read' task assesses students' pronunciation, which is a part of speaking skills; the 'listen and choose' task requires students to listen for the target information and provide a formatted spoken answer; the 'answer questions' task takes a more open format and assesses students' ability to listen for information and answer the questions more independently; the 'retell a story' task assesses students' skills to take notes on a short story and narrate the story based on their notes; the 'ask questions' task assesses students' ability to ask questions according to the task requirements, attempting to add a more communicative colour to the test. In general, the listening-and-speaking test adopts a rather integrative test design, which seeks to assess different aspects of communication skills, such as pronunciation, listening for details, listening for gist, and responding according to the context. From my perspective, it goes further than the textbook's design on listening and speaking skills, as the textbook arranges separate sections for these two skills and rarely offers opportunities to practice these two skills as a whole.

4.2.2 Written test

The written test involves four parts: a cloze for vocabulary, reading comprehension, a cloze for grammar, and writing. The two cloze tasks, each worth ten points, adopt different designs. The cloze for vocabulary (see Appendix 12, p. 282) is a multiple-choice task. An around-220-word passage in which ten words or phrases are omitted from the text is presented to students. For each gap, four choices are provided for students to choose from, and only one answer is correct. The task aims to examine students' "knowledge of vocabulary and collocations" and requires them to "understand the words' meanings in the context" (Ministry of Education, 2011, p. 19). On the other hand, the cloze for grammar (see Appendix 12, p. 288) is a gap-filling task. Another around-220-word passage with ten words or phrases omitted is displayed for students. For some gaps, hints, usually a different form of the omitted words (e.g., value – valuable), are provided in brackets. For other gaps, in which the omitted word is a pronoun, a conjunction, a modal verb, or an article, no hint is provided. Students are required to fill in the missing words by analysing which part of speech is missing. Their ability to "use their grammar knowledge" is assessed in this task (Ministry of Education, 2011, p. 19).

The reading comprehension (see Appendix 12, pp. 283-288), worth 40 points in total, includes five reading tasks. The first three tasks require students to read three passages, which are around 250 to 350 words, and answer five multiple-choice questions (each worth two points) for each passage. The questions are designed to examine whether students can "use reading strategies to obtain information" from the text (Ministry of Education, 2011, p. 17). They focus on the "details" and the "main idea" of the passages and the "logical relationship" between different sentences to see if students can identify the specified information and unpack what the author intends to convey (Ministry of Education, 2011, p. 17). The latter two tasks are matching tasks, which involve ten sub-tasks worth ten points in total. The first task requires students to match five sentences out of six choices into a 250-word passage, in which five sentences have been excluded, while the second task asks students to match five pieces of text out of six to five sentences. These tasks examine students' ability to infer what idea the author may communicate from the logical relationship of the context and identify information from the sentences to recognise connections between different texts. They assess students' abilities to "understand various details" through reading, which is a skill enlisted in the level descriptors (Ministry of Education, 2011, p. 17).

The writing task (see Appendix 12, p. 288), worth 15 points, requires students to write an 80-word article, which could be a report, an email, or a note, according to a given context. The task instructs students about the key ideas that should be addressed in the article in Chinese and offers the English translation of some keywords that might be challenging to the students. A sentence to begin with in writing is supplied. According to the level descriptors in the curriculum, students' writing abilities can be described from task achievement, cohesion, and coherence perspectives. According to the curriculum, a middle school graduate should be able to "write short articles independently" "according to the requirements of the task", "use proper conjunctions to indicate the logical relationships between sentences", and "describe people and events in a clear way" (Ministry of Education, 2011, p. 17).

The task design also shows that translation seems to be an important aspect of students' capacity examined by the writing task. Since the task requires students to include the given main points in their writing, and the main points are addressed in Chinese instructions, the students need to translate the instructions to construct their articles. However, translation is not mentioned in the curriculum as a skill that is included in writing or a skill that students

should master. It cannot be taken as the same as writing, as translation largely requires fixed responses and is a much less creative process than writing (Min et al., 2020). Therefore, a presumption is made that the Chinese instructions may be an approach of scaffolding to enable students to write in a more organised way. Another presumption is that such a design may serve the purpose of fixing the format and content of students' responses, which can make the marking job easier.

Overall, the written test assesses vocabulary, grammar, reading, and writing through different tasks. The cloze tasks and reading comprehension tasks assess students' mastery of vocabulary and grammar and their skills to read texts with moderate length. All of these tasks are closed tasks, including multiple-choice and gap-filling. They match the textbook design, in which vocabulary, grammar, and reading are given excessive attention. The writing task, on the other hand, assesses translation skill, a skill not addressed by the textbook or the curriculum, along with writing skills. This might be a design not for fostering better skills of students but for convenience reasons, both in teaching and administration.

4.3 Summary

This chapter reviews the designs of the textbook, guidebook, and English Zhongkao that impacted the context under investigation. Based on the official curriculum, the textbook and guidebook follow a design that intends to strike a balance between teaching language knowledge and teaching communication. While language knowledge such as vocabulary and grammar are emphasised in teaching, the textbook and guidebook also stress the importance of using such knowledge in different settings and recommend activities that facilitate students in practising the knowledge for communication. The textbook divides the units into different sections, each focusing on a particular language knowledge or language skill. The design appears to consider these skills and knowledge as discrete and places more emphasis on reading skills, vocabulary, and grammar. The guidebook encourages teachers to adjust their teaching based on the local context instead of ordering them to perform specific actions. The suggestions made, however, mainly concentrated on teachers' pedagogical activities rather than assessment activities and offered little instructions on whether and how teachers should provide feedback to students.

The English Zhongkao adopts a design that examines students' language abilities integratively. The listening-and-speaking test is an integrative test requiring students to combine their listening and speaking skills to complete the tasks (Hughes, 2003). The test reflects the curriculum's descriptions of middle school graduates' listening and speaking abilities, emphasising their communicative skills and effectiveness. The two cloze tasks and the reading comprehension tasks can also be classified into integrative tasks, as they integrate different areas of language knowledge within a textual context where reading skills are required (McNamara, 2013). It should be noted that these tasks are worth 60 points in total and take a larger share of the total scores than the listening, speaking, and writing parts combined. It seems that the textbook, guidebook, and English Zhongkao reach an agreement that reading skills, vocabulary, and grammar deserve more attention in teaching and testing. The writing task uses Chinese, the first language of the students, to write instructions and requires students to include the main ideas provided in their writing, which makes the writing task assess writing as translation, despite that translation is not a skill required by the curriculum. The task design may serve purposes other than assessing students' language skills, including assisting students in writing or making marking tasks easier.

Chapter 5 Case study one: School A

Case study one adopts both data-driven inductive and theory-driven deductive approaches to analyse teachers' assessment practices and their understanding of assessment and assessment policies within the setting of School A. It starts with a profile of School A (section 5.1), introducing the pertinent information to the study of the two teacher participants, T1 and T2, and School A, the context in which the teachers worked. The teachers' descriptions of the students' language level and the sitting arrangements of the classrooms were also presented to offer an overview of the classroom contexts. The following sections document the assessment activities implemented in T1's (section 5.2) and T2's (section 5.3) classrooms. Section 5.4 details the teachers' understandings of assessment, and section 5.5 reports on their understanding and enactment of the assessment policies that impacted their working environment. Section 5.6 summarises the findings of this case.

5.1 Profile

T1 was an experienced teacher who had teaching experience for over 17 years. Majored in English education, T1 received professional education in English language teaching and learning at a local teacher training college in her hometown. She took the role of a high school English teacher in her hometown after graduation and became a middle school English teacher when she moved to Shenzhen. T2 was an early-career teacher who had only started her second year of teaching when the study was conducted. As a recent graduate who only obtained her bachelor's degree a year ago, T2 had limited academic or professional experience in English education. She earned her bachelor's degree at a local university in Shenzhen, studying English-French bilingual major, which was a language-focused programme and did not include education-related courses. Her most relevant experience in teacher education during her university life was obtaining a teaching certificate before graduation and being an intern teacher in a middle school in Shenzhen for a month.

School A was a public secondary school that provided both middle school education and high school education. It had an enrolment of over 1700 students in the middle school department. Each class in the middle school department had around 48 students. Each English teacher in the middle school department was responsible for two classes, around 96 students in total. As stated by T1 and T2, School A had the following features in English

language education. Firstly, the teachers argued that the overall students' English level was below average compared to other public middle schools in the district. Both teachers referred to exam results as the evidence of students' level of English and believed there was a gap between students' language level in School A and students in other public middle schools:

> In the unified examination (the exam all students in the district take at the end of a semester), our average score of English is always five to eight points lower than the average (score of the middle schools in the district). (T1:1)

> We had an exam at the beginning of the first semester of grade 7 – the total score was 40 points, and the average score was only 20 … it's very low … comparing to students in other schools. (T2:2)

T2 also referred to how her students responded to her teaching as evidence of students' language level. She reported that although the local education bureau encouraged teachers to speak English in class, many of her students "have trouble learning through English instructions" (T2:1).

The second feature reported by the teachers was the "polarisation" in the students' performances in both classrooms and tests (T1:1; T2:1). As T1 suggested, there was a "significant disproportion between the number of high-performing and low-performing students" within her class and the school (T1:1). For example, in her class, "only two to three students could achieve high scores, and many students could get no more than 30 points out of 100" (T1:1). Again, she resorted to test results as evidence of students' performance. T2 agreed with T1's report regarding the sharp difference between the English level of different groups of students. According to T2, while a few students in her class could respond well in classroom interactions and perform well in tests, there remained a large group of students who "could not keep up with their higher-performing peers" and "lack the enthusiasm to work harder" (T2:1-2).

The teachers also reported on how the classroom layout is arranged in the two classes. In both classes, students sat in rows, facing the screen, the blackboard, and the teacher's desk at the front. The classroom layout in T1's class was decided by the class teacher (班主任) of the class, who oversaw student administration, and T1 helped liaise with the class teacher to explain the sitting arrangements. The class teacher of T1's class decided where students sat based on two factors: "the height of students" and "student discipline" (T1:informal

interview). Taller students would usually be arranged to sit at the back, and students who were considered constant misbehavers would be arranged to sit away from those who were also considered misbehaving. The class teacher of T1's class would take students' preferences for sitting arrangements into consideration. However, she did not follow whatever the students asked, as she believed students preferred to sit with their friends, and "sitting with friends could potentially damage classroom discipline and the efficiency of teaching and learning" (T1:informal interview). T2 was the class teacher of her class and was responsible for the sitting arrangement. She regarded student discipline highly and would "ask the misbehaving students to sit at the two sides of the class so that their bad influences on other students' learning can be minimised" (T2:informal interview). Other than that, the students were randomly allocated. If the students were not satisfied with the arrangement, they could negotiate with other students for a possible alternative seat, but T2 was the one who made the final decision.

From the teachers' report, the sitting arrangement was a way of managing classrooms in the two classes. The teachers considered discipline as a crucial factor for success in learning. Students' preferences were taken into account but were not considered essential. Although such approaches could result in a less collaborative relationship between students in classroom learning, it cannot be denied that teachers adopt such arrangements with the hope that a more disciplined classroom could contribute to a better learning environment. It should also be borne in mind that such arrangements were made for classes of around 50 students. The large class size teachers coped with should be considered while analysing teachers' assessment activities and understandings in the following sections.

5.2 T1

This section reports on T1's teaching and assessment in the classroom. Section 5.2.1 presents T1's intended teaching objectives of the unit and reviews her actual teaching arrangements, which illustrate the background of her assessment designs. Section 5.2.2 describes and examines the observed assessment activities in T1's class.

5.2.1 Background of assessment design

The lessons observed in School A were based on unit 2 of the textbook, around the theme 'numbers'. Before observing her classroom assessment activities, T1 was interviewed

regarding her intended teaching objectives of the observed unit. From her perspective, the objectives of the unit were three-fold. The first objective she mentioned was vocabulary. T1 believed that students should "learn the pronunciation, meaning, and usage of the new words, and use them correctly" (T1:16). The second objective was understanding the reading passages. T1 considered that students should "remember the keywords and phrases in the passages learnt, understand the structures of the sentences, and recite and dictate the selected texts" (T1:16). The third objective was grammar. Since this unit's grammar knowledge was cardinal number and ordinal number, T1 argued that "students should learn what is cardinal number and ordinal number, understand when to use these numbers, and be able to finish exercises related to these numbers" through learning this unit (T1:16). She appeared to consider linguistic knowledge and reading as a discrete language skill as the priority of her teaching. This suggests that she viewed language as a set of linguistic elements and separate skills which needed to be mastered independently, which resonated with the behaviourist view of language learning (Brown & Abeywickrama, 2019).

As T1 argued, the reason why more attention was drawn to the practice of vocabulary, grammar, and reading skills was that the formal examinations they took, including unit test, final test, and Zhongkao, "placed a high emphasis on reading comprehension and the mastery of vocabulary and grammar" (T1:16). According to T1:

> The Zhongkao our students take … has two cloze tasks, and the score for reading comprehension has recently been increased to 40% ... (students) need to be skilled in vocabulary and grammar and competent in reading to succeed in Zhongkao. (T1:2-3)

T1 mentioned the cloze tasks, which examined students' "reading skills and command of language knowledge, such as grammar and vocabulary", and the reading comprehension tasks, which focused on students' "reading comprehension skills", as the reasons for her choices of teaching objectives (Shenzhen Admission and Examination Office & Shenzhen Institute of Education Sciences, 2021, p. 69). From her perspective, it was necessary for teachers to "spend more time teaching knowledge that is most needed in the tests" and "assign more exercises for students to practise the target skills" (T1:2-3). Surprisingly, other language skills articulated in the curriculum, such as speaking, listening, and writing, were not mentioned by T1 as the objectives of this unit, although they were also tested in Zhongkao. The large share of scores reading comprehension and cloze take in the English

Zhongkao might make T1 feel that skills and knowledge tested in these tasks are more important than others.

As for the teaching arrangements of the unit, T1 argued that she referred to the official textbooks, the corresponding teacher guidebook, and workbooks as her primary lesson planning resources and would usually "follow the arrangement of the textbook unit" by teaching "the vocabulary list first, the reading part second, including the more practice part ... then the grammar, the listening and speaking parts, and the writing" (T1:3). Among the seven observed lessons, T1 assigned three lessons for teaching reading passages and explaining reading exercises, three for teaching vocabulary, grammar and explaining related exercises, and one for listening and speaking. However, the observation showed that T1 skipped several unit sections: the speaking section was replaced by the listening-and-speaking mock test to "examine students' listening and speaking skills" (T1:12); the 'getting ready' and 'self-assessment' sections, the assessment pack that encourages students to review their progress (discussed in section 4.1.1, p. 73-74), were excluded because she "seldom list the lesson aims … and having students tick the checklist was of little value to teachers' teaching" (T1:23); the writing section, which required students to write a short article with numbers, was omitted because the task was "too easy and will not be tested in exams" (T1:informal interview); the project section, which allowed students to collaborate on tasks related to the unit's topic and demonstrate the knowledge and skills they learned, was also ignored, as this section "was too difficult for students and cost too much time" (T1:informal interview). It seemed that T1 placed her emphasis mainly on teaching the unit's vocabulary, grammar, and reading section, which matched her narratives about her intended teaching objectives, and selected textbook parts into her teaching plan based on their relations to test. Test appears to be a significant factor that directs her teaching and assessment.

5.2.2 T1's assessment activities

Through the classroom observations, T1's classroom assessment activities were documented and summarised. Among the seven lessons observed in T1's classroom, 60 assessment activities were identified and classified into three types – oral assessment activities (section 5.2.2.1), written assessment activities (section 5.2.2.2), and student-assessed activities (section 5.2.2.3). The percentage of each type of activity was calculated and reported in Table 5-1.

Assessment activity	Times	Assessment activity	Times	Assessment activity	Times
TO-V	17 (28.3%)	TW-RE	9 (15%)	PW-RE	1 (1.7%)
TO-RC	11 (18.3%)	TW-GE	4 (6.7%)	SW-LE	1 (1.7%)
TO-G	5 (8.3%)	TW-LE	3 (5%)		
TO-T	4 (6.7%)	TW-D	1 (1.7%)		
TO-RA	4 (6.7%)				
Teacher oral total: 41 (68.3%)		Teacher written total: 17 (28.3%)		student-assessed total: 2 (3.4%)	
Total: 60					

Table 5-1 T1's assessment activities

5.2.2.1 Oral assessment activities

The observation showed that T1 tended to assess students through oral assessment tasks, as 68.3% of the assessment activities were employed to elicit students' oral responses. T1 regularly raised planned or improvised questions to students based on the textbook content. She sometimes invited individual students to answer the questions and sometimes raised the questions to the whole class without assigning the questions to specific students. These questions were mostly display questions (Walsh, 2011), where teachers already knew the answers (for example, 'what does the word palace mean?') and required students to display what they knew. They generally took closed formats and aimed to elicit restricted responses from students to check their understanding. Among these questions, vocabulary, reading comprehension, and grammar appeared to be the three primary assessed constructs in T1's classroom assessment activities, which matched with T1's reported teaching objectives. For vocabulary, T1 usually examined students' recall of the pronunciation and meaning of the words and phrases learnt in the unit. For example:

> T1 gives the Chinese meaning of the important phrases in the passage, asks students to answer the corresponding English phrases, and then requires them to find these phrases in the passage and highlight them with a red pen. (T1:L2)

> T1 asks students, "do you remember the word 'abacus'" (students learnt this word yesterday). Students answer the Chinese meaning of the word together, and T1 confirms their answers. (T1:L8)

For reading comprehension, T1 checked students' reading skills, such as skimming and scanning, and understanding of the details of passages in the textbook, workbooks, and exercises. For example:

> T1 asks one student to answer a question related to the textbook passage, "what was the king's problem". The student cannot answer immediately. T1 asks, "where can you find the answer in the passage". The student scans the passage and finally answers, "he did not have enough rice for all the squares." T1 confirms his answer. (T1:L2)

As for grammar, T1 raised examples or counterexamples to check whether students understood the grammar points of the unit. Translation tasks were also assigned to evaluate students' memory of vocabulary and use of grammar. For example:

> T1 asks, "should we add '-s' to 'two hundred'". Some students say yes, but some remain silent, so T1 emphasises there is no need to add "-s" to hundred, thousand, million and billion while reading specific numbers unless they use phrases such as "hundreds of", "thousands of", "millions of". (T1:L3)

> T1 shows example sentences "The number of students is two hundred" on the screen and asks one student to translate the sentence into Chinese. After the student answers, T1 explains the phrase "the number of" and why "is" is used in the sentence. (T1:L1)

Besides vocabulary, reading, and grammar, T1 also assessed students' reading aloud through oral tasks. For example:

> T1 asks students to read the passage after the recording (one sentence at a time). T1 corrects students' pronunciation. (T1:L2)

After receiving students' responses, T1 provided feedback regarding students' performances. There were mainly three types of feedback offered by T1. The first type, confirmatory feedback, a type of positive feedback (Hattie & Timperley, 2007), was the most often observed feedback in T1's class, particularly while T1 raised display questions and expected students to demonstrate what they knew. If the students responded to her with the anticipated answers, T1 confirmed students' answers by saying, "Yes, very good" or "Ok, sit down please". The first reason why confirmatory feedback was implemented the most often might be partly ascribed to the closed formats of T1's questions, in which a limited amount of information could be collected from the responses elicited, and spaces for further feedback remained constrained. Such an assessment approach could be viewed as repeated summative assessment, in which judgmental evaluation outweighed qualitative advice (Torrance & Pryor, 2001). A second reason might be attributed to the classroom climate. According to the classroom observation, when T1 raised a question to the whole class, it was usually the case that a small number of high-performing students actively engaged with T1's questions and

could often answer the questions correctly, while the rest of the class remained quiet and only took part in the classroom conversations when invited. This might result from the shape difference between the language level of different groups of students, as T1 described in section 5.1. The confirmatory feedback was, therefore, often provided to the active students. Other students who tended to be passive during the oral assessment activities had limited opportunity to receive feedback unless they were nominated to answer the questions.

The second type, corrective feedback, "the responses to a learner's nontargetlike L2 production" (Li, 2010, p. 309), was also identified in T1's class, notably when T1 corrected students' pronunciation during reading aloud activities. When T1 noticed students' pronunciation mistakes while reading the textbook passages, she pointed out the mistakes, read the words or sentences for students, and asked them to repeat after her. For instance:

> The students read the passage after the recording and made mistakes in their pronunciation: "One day, a wise old man came to the place [palace]." The students were having trouble pronouncing the word "palace", so T1 read the word "palace" to students, and students read after her. (T1:L2)

The third type of feedback adopted by T1 was sharing success criteria with students. T1 informed students of goals related to their performance regarding how she wanted them to answer her questions or finish the exercise tasks. A representative example was the 'full-sentence criterion' T1 constantly articulated in class, asking students to answer her questions in a complete sentence instead of with single words or phrases. For example:

> T1 asks one student, "what was the king's favourite game?" The student answers, "chess". T1 says, "can you answer in a full sentence?" So, the student says, "the king's favourite game was chess". (T1:L2)

T1 suggested that it was necessary to remind students about answering in a complete sentence because "students can build a sense of complete sentences and practice speaking more" (T1:22). She also argued that this was "a preparation for the listening-and-speaking test in Zhongkao" because "the test requires students to answer in full sentences" (T1:22). It seemed that tests defined the criteria T1 shared.

From the observation data, T1's feedback mainly focused on positive evaluations of students and whether a target response was produced but hardly addressed students' learning process and what they should do to make further progress. The questions T1 raised might be one of

the reasons for such a situation. According to the data, most of T1's oral assessment tasks were closed questions concerning specific vocabulary, grammar, and details in reading passages that were intended to draw out fixed responses from students. The limited space for answering restricted students' demonstration of their capacity and offered reduced room for T1 to comment on students' performance. Such a situation was aggravated by the passive classroom climate, in which many students were reluctant to speak up, and classroom interactions became a game between the teacher and the active students.

5.2.2.2 Written assessment activities

The observational data showed that T1 adopted 28.3% of the assessment activities to elicit students' written responses. These activities included one dictation and various exercises for reading comprehension, grammar, and listening and were adopted during lessons to assess students. The dictation was adopted to assess and reinforce students' memorisation of vocabulary. At the same time, the exercises took the form of cloze, multiple-choice questions, and short-answer questions, which were fashioned after the test items in Zhongkao, and assessed language knowledge and skills, such as grammar, reading, and listening skills. The following task excerpts show examples of different written exercises, including cloze, reading exercise, and listening exercise, assigned in T1's classroom.

Task 1

Read the textbook passage and complete the cloze task.
A long time ago, there was a king who 1. _____ (live) in India. 2. _____ (play) chess was his favourite hobby. One day he challenged a wise old man 3._____ a chess game.

Task 2

Read the passage and choose the answers.
1. What do we know about diet soft drink? ( )
A. They don't have any sugar or calories. B. They can help people lose weight.
C. They don't taste sweet. D. They aren't good for our health.

Task 3
Listen to the dialogue and answer the following questions.
1. When will the woman go to see the doctor? (10:30am/3:00pm/3:30pm)

_____

These written tasks were mainly exercises in the textbook and the workbook subscribed to students. They were generally designed according to the textbook content, which addressed the language knowledge (such as vocabulary and grammar), the topic of the unit (in this case, number), and other language skills that needed to be practised (such as reading and listening). Therefore, T1 selected exercises according to the focus of her lessons and assigned these tasks as in-class exercises or after-class homework to check whether students had mastered the target knowledge and skills. If students completed the tasks in class, T1 invited them to share their answers and provided feedback based on these answers. If the tasks were assigned as homework, T1 collected and marked the worksheets after students finished and later explained the tasks in class. The interactional process involved during the exercise explanation resembled those presented during the oral assessment tasks. For example:

> T1 reads the reading comprehension question in the workbook, "why is Longjing tea leaves famous", and invites a student to answer. The answer can be located in the second paragraph, so the student reads from the paragraph, and T1 confirms the answer. (T1:L4)

Feedback included during T1's written assessment activities concentrated mainly on the task level, which resembled those involved in oral assessment activities. Confirmatory feedback was provided if students arrived at the right answer; success criteria were also shared for specific tasks, such as the listening-and-speaking test in Zhongkao, to help students familiarise themselves with the task type so that better test-taking skills could be developed; corrective feedback was generally supplied if students failed to complete the written tasks accurately. Further elaborations on the written tasks, particularly translation of key information related to the answer keys, were provided if needed to assist students' understandings of the tasks. For example:

> T1 explains the reading exercises in the workbook. She translates the first question and asks students which choice is the right answer. Students respond with different choices, so T1 translates the paragraph related to the question to explain the correct answer. (T1:L4)

5.2.2.3 Student-assessed activities

Apart from raising questions and assigning written tasks, T1 also sought to engage students in assessment activities by having them work individually or in pairs on rare occasions. Two activities involving students as assessors, which accounted for 3.4% of the observed

assessment activities, were observed in T1's class. Both activities focused on students' written performance in exercises and were carefully controlled by the teacher. The first activity took place while T1 instructed students to check their peers' answers before she explained a reading comprehension exercise:

> T1 asks four students to write their answers on the blackboard. Then, T1 asks another four students to correct the answers on the board if they have different answers. T1 then marks the answers on the blackboard and explains each task. (T1:L4)

The second activity happened after T1 instructed students to finish a listening exercise. After students finished the task, T1 displayed the answer keys on the screen and required students to self-evaluate the answers:

> T1 plays the recording, asking students to do the exercise in the textbook (p22). After finishing, T1 plays the recording again and shows the answers on the screen. T1 asks students to self-check the answers. She only explains some of the tasks without checking students' answers. (T1:L3)

T1 did not provide any feedback to students during this activity. She explained that this listening exercise was "quite simple" and did "not need detailed explanations" (T1:informal interview). While T1 described these activities in the after-observation interview as "self- and peer assessment" (T1:19), from these two activities, it seemed that students were mainly involved through self- and peer-marking. Little information was collected during these activities to inform T1's teaching, and the feedback students received was confined to the answer keys and the explanation of the exercises.

5.3 T2

This section details T2's teaching and assessment in the classroom. Section 3.1 illustrates T2's reported teaching objectives of the unit and her actual teaching arrangements, which exhibit the background of her assessment designs. Section 3.2 demonstrates and analyses T2's classroom assessment activities.

5.3.1 Background of assessment design

T2 reported that she referred to three resources for her teaching and assessment planning: well-designed teaching slides for the textbook, workbook, and experienced colleagues'

suggestions. From T2's perspective, as a novice teacher with limited teaching experience, she could learn from these resources about "which parts of the textbook are more often explained and examined" (T2:3). T2 did not mention the teacher guidebook as a helpful resource. She argued that although she tried to follow the suggested teaching design offered by the guidebook, she found that such design was not "sufficiently effective in terms of improving students' test performance" or "fostering students' language skills" (T2:3-4). T2 attributed such a phenomenon to students' poor language level:

> Those interactive activities made the classroom very lively, but if my students did not have enough language knowledge, they couldn't learn much from these activities … they were just having fun. (T2:4)

As a result, T2 believed that "a pragmatic approach" was what she should choose (T2:4). She reported that such an approach had brought much better outcomes:

> Now, I prefer teaching the vocabulary first, instructing students to learn the pronunciation and other related words, and then teaching the reading passages, picking up the "real stuff", you know, vocabulary, grammar, and key sentences, and asking them to recite, write from memory and do more exercises … Now, they can at least use some of the words and sentences we learnt. I think this is much more effective (T2:4)

T2's reported teaching objectives of the observed unit corresponded with her 'pragmatic approach'. She considered that the objectives of the unit were two-fold. The first one was learning numbers, which was the unit's topic. T2 wanted her students to "review the usage of cardinal numbers and ordinal numbers", which were the grammar points of this unit (T2:13). The second one was understanding the reading passages in the unit. T2 believed, through learning the passages, students should be able to "improve their language ability, including their command of vocabulary, their translation ability, and their understanding of longer texts" (T2:13). Interestingly, other language skills specified in the textbooks, including speaking, listening, and writing, were not mentioned by T2 as a part of her intended objectives or 'pragmatic approach'. It seemed that, similar to T1, T2 considered language learning as acquiring separate linguistic components and reading skills, which suggests that she adopted behaviourist perspective in language teaching (Brown & Abeywickrama, 2019) and tended to focus on the textbook sections that were more related to students' test-taking in her teaching.

The observation showed that T2 has mainly concentrated on teaching vocabulary, grammar, and reading in this unit. In her eight lessons, T2 allocated three lessons for reading passages in the textbook, four for vocabulary, grammar, and related dictations and exercises, and one for a listening-and-speaking mock test. Similar to T1, T2 skipped several sections of the unit: the speaking section of the unit was substituted with a listening-and-speaking mock test for "test preparation" (T2:7); the 'getting ready' and 'self-assessment' sections were ignored because "the checklists were of little use" (T2:19); the writing section was excluded because "the writing tasks (of this unit) will not be tested" (T2:informal interview); the project section was also omitted, as this section "was usually replaced with exercises from the workbook since the exercises address similar knowledge and are more relevant to students' test-taking" (T2:informal interview). Relevance to test items in Zhongkao appeared to be T2's criteria for selecting content for her teaching.

5.3.2 T2's assessment activities

T2's classroom assessment activities were summarised and documented through classroom observations. Among the eight lessons observed in T2's classroom, 66 assessment activities were identified and classified into three types – oral assessment activities (section 3.2.1), written assessment activities (section 3.2.2), and student-assessed activities (section 3.2.3). The percentage of each type of activity was calculated and reported in Table 5-2.

Assessment activity	Times	Assessment activity	Times	Assessment activity	Times
TO-G	16 (24.3%)	TW-RE	5 (7.6%)	PO-G	4 (6.1%)
TO-V	14 (21.2%)	TW-GE	5 (7.6%)	PO-S	1 (1.5%)
TO-T	6 (9.1%)	TW-LE	3 (4.5%)		
TO-RC	3 (4.5%)	TW-D	3 (4.5%)		
TO-R	2 (3%)	TW-WFM	3 (4.5%)		
TO-S	1 (1.5%)				
TO total: 42 (63.6%)		TW total: 19 (28.8%)		PO total: 5 (7.6%)	
					Total: 66

Table 5-2 T2's assessment activities

5.3.2.1 Oral assessment activities

The observation showed that T2 tended to implement oral assessment tasks more, as 63.6% of the assessment activities targeted students' oral responses. Like T1, T2 generally raised the questions to either individual students or the whole class. The questions T2 raised were

generally display questions based on the textbook content. Grammar and vocabulary seemed to be the two major themes in T2's oral assessment tasks, which corresponded with her intended teaching objectives. For grammar, T2 usually checked students' understanding of the target grammar with counterexamples. Translation tasks were also adopted to examine students' understanding of grammar usage. For example:

> T2 asks, "can I say, 'three hundreds and fifty-five'?" Students say, "no!" T2 asks, "why?" Students explain, "you shouldn't add '-s' to 'hundred'". (T2:L4)

> T2 provides the Chinese sentence "今天天气真热。你不想喝一杯可乐吗？(It is such a hot day. Wouldn't you like to drink a cup of cola?)" and asks one student to translate the sentence to check student's understanding of the usage of 'Wouldn't you …" (T2:L3)

As for vocabulary, T2 regularly raised questions to examine students' memory of the pronunciation, spelling, and meaning of words and phrases. Translation tasks were also applied to assess students' usage of particular words. For example:

> T2 reads the third sentence in the task and asks students, "what is 'challenge'?" Students translate the word into Chinese. (T2:L2)

> T2 nominates one student to spell the word "parents". (T2:L8)

> After explaining the word "control", T2 asks students to translate a Chinese sentence "不用担心，一切都在控制当中 (Don't worry. Everything is under control.)"into English. (T2:L8)

Other aspects of language learning were evaluated less frequently through oral assessment tasks in T2's class. Three activities were identified to evaluate students' reading skills and understanding of textbook passages; two required students to recite selected text from the textbook; one storytelling activity was arranged to encourage students to speak. For instance:

> T2 asks students to read paragraph one and two and answer the question, "where did the story happen". (T2:L2)

> In the grammar lesson, after introducing the ordinal numbers "first, second, third", T2 asks students to recite a paragraph from the textbook passage which contains the ordinal number "second". (T2:L4)

> T2 invites students to tell traditional Chinese stories about numbers. (T2:L2)

Similar to T1, T2 implemented three types of feedback during oral assessment activities. The first type was confirmatory feedback, which accounted for the largest segment of T2's responses to students' performance. T2 typically confirmed students' answers by saying "yes, [repeat the answer]" or moving forward without saying anything. Repeating the answer represents the recognition from T2, while not saying anything, according to T2, was more like an "implicit understanding" between T2 and the students: "if I do not say anything, there is nothing wrong with the answer" (T2:17). The closed format of T2's questions might be a reason for the prevalence of such restricted responses in her class. Like T1, T2 adopted continuous questioning to evaluate students' understandings of the learnt knowledge. Whether students could produce an expected, correct answer appeared to be the focus of her assessment. A second reason might be related to the students' characteristics. As T2 argued, many students in her class were "shy" and would not respond to her questions "unless they were certain that their answer was correct" (T2:10). In T2's opinion, "comparing to picking students up on their mistakes", it was "more important to protect their self-esteem and encourage them more" (T2:10). Such a position might explain why confirmatory feedback was widely employed in T2's class. A third reason might be that T2 intended to keep the difficulty level of her questions moderate so that most students could manage to answer properly. Such an approach might maintain students' confidence and motivation to respond to teachers.

The second type was sharing success criteria with students. Like T1, T2 often reminded her students of the 'full-sentence criterion', that is, answering questions in a complete sentence instead of in single words or phrases. T2 explained that "my students like to answer questions with a single word or phrase; however, if they do that in the listening-and-speaking test, they may receive low scores" (T2:10). She also argued that asking students to speak in complete sentences allows her better opportunities to assess students' answers: "If their answers were too simple, it would be difficult to identify their problems" (T2:16). However, T2 also suggested that if low-performing students had trouble producing complete sentences and could only respond with single words, she would "accept the answers as long as they addressed the key points" and "reiterate the answer in a full sentence", as "every progress these students made should be recognised" (T2:16).

The third type was disconfirmation, a type of negative feedback (Hattie & Timperley, 2007), with additional questions or comments. If students provided a wrong answer, T2 asked

additional questions such as “is it correct” or “are you sure” to suggest that the given answer might not be correct and imply that students should give it another try. Besides additional questions, T2 also provided comments to indicate whether the students’ answers met her expectations. For example:

> T2 asks, “besides ‘spend time doing something’, what else can we say with ‘spend’?” One student answers, “spend on do something”. T2 asks, “are you sure?” The student hesitates, “to do something …” Other students correct him, “spend on something”, so the student finally says, “spend on something”. T2 confirms the answer by repeating “spend on something” and writing the phrase on the blackboard. (T2:L4)

> T2 invites one student to complete the second blank of the table. The student says, “the king and an old man”. T2 asks the students to answer in a complete sentence, and the student says, “the king and an old man in the story”. T2 asks the student to sit down and invites another student to answer. The student says, “there are two people, the king and a wise old man, in the story”. T2 comments, “it’s a little strange”, and asks a third student to answer in a simpler way. The third student says, “the king and a wise old man was in the story”. T2 asks, “should you use ‘was’?” Other students say, “were!” Finally, T2 shows the correct answer on the screen, “the king and a wise old man were in the story.” (T2:L2)

The observation showed that T2 tended to construct more complex classroom conversations with questions and comments to students. Potentially, such feedback could engage more students in oral assessment activities. However, it could not be overlooked that T2 placed most of her emphasis on how well the tasks were accomplished and error correction at the task level but hardly addressed the thinking process required for similar tasks and how better progress could be made. The closed questions T2 raised might also restrain her feedback’s effectiveness, as students merely demonstrated if they understood the fragmentary linguistic knowledge and accomplished tasks with it instead of what they understood and could do with the knowledge, which made it problematic for T2 to provide instructions for future development.

5.3.2.2 Written assessment activities

The observational data showed that 28.8% of T2’s assessment activities were associated with written tasks, such as dictations, write-from-memory tasks, and exercises. Dictations and write-from-memory tasks were adopted multiple times to assess students’ recall of key words and phrases in the unit and selected text from textbook passages. The exercises T2 chose, on

the other hand, were mainly from the textbook and the workbook prescribed for students. The tasks addressed the topic and language knowledge of the unit and focused on language skills, such as reading and listening, and took the form of cloze, multiple-choice questions, and short-answer questions. Some of these tasks were mock tests fashioned after the test design of Zhongkao. The following task excerpts show examples of written exercises, including reading, grammar, and listening exercises, adopted in T2's classroom.

Task 1
Read the passage and choose the answers.
1. What do students need to do when they visit Cadbury World? ( )
A. To go with their parents. B. To bring their ID cards.
C. To wear uniforms. D. To provide student cards.

Task 2
Cloze.
Before the invention of written numbers, people used many 1. _____ (difference) ways to count things. First, people used 2. _____ (they) fingers, and even their toes. 3. _____, they could only count small numbers in this way.

Task 3
Listen to the dialogue and answer the following questions.
5. Who will go and buy some drinks (The boy's mum/The boy/Mary)
_____

Similar to T1, T2 also assigned these written tasks as in-class exercises or after-class homework based on the focus of her lessons. If the tasks were distributed as homework, T2 marked the worksheets to overview students' performance and address the problems she discovered in class. If the tasks were assigned as in-class exercises, she invited students to share their answers and then provide feedback or explain the tasks. The classroom conversations around these exercises resembled the dialogue patterns that occurred during the oral assessment tasks. For example:

> T2 invites one student to complete the table in the workbook. The student first says, "it take place in India." T2 says, "it …", hinting there is something wrong, and the student quickly changes the answer to "it took place in India". T2 confirms the answer. (T2:L2)

The feedback T2 incorporated in written assessment activities mainly related to the task level, which bore a resemblance to the feedback she applied in oral assessment activities.

Confirmatory feedback was given when students' answers were accurate; the 'full-sentence criterion' was emphasised when students responded to short-answer questions; disconfirmation with additional questions or comments was generally supplied if there were problems in students' answers. Further elaborations were also provided, typically through translating the tasks' text, to assist students' comprehension of the tasks and the related language knowledge. For instance:

> T2 starts explaining the first cloze. She first waits for the students to share their answers altogether. If students responded with wrong answers, T2 translated the corresponding text to explain which answer should be chosen. (T2:L3)

5.3.2.3 Student-assessed activities

Besides teacher-assessed activities, five student-assessed activities were observed in T2's class, which accounted for 7.6% of the observed assessment activities, a slightly higher figure compared with those in T1's class. Students acted as the assessors of their peers' oral performance in these five activities. Three activities took place while T2 instructed students to assess whether their peers' answers were correct. For example:

> T2 asks one student to translate the Chinese phrase "花时间在某件事上 (spend time on something)". The student says, "spend time do something". T2 repeats his answer and asks other students, "is it correct?" Other students correct the answer, "spend time on something". (T2:L4)

Another two peer assessment activities happened without T2's instructions. The observation revealed that students sometimes assessed their peers' answers autonomously, especially when their peers had trouble answering a question or completing a task correctly. For instance:

> T2 invites four students to write designated cardinal numbers in English on the blackboard and asks other students to write the numbers on their books. After the four students finish writing, T2 checks and corrects their answers on the blackboard, while other students respond to the answers with "right", "wrong", and how the numbers should be spelt. (T2:L4)

Among these activities, students engaged in error detection and correction with or without T2's supervision. They acted as the chief feedback providers who supplied confirmation, disconfirmation, and corrective feedback. T2 reported in the after-observation interview that "students sometimes enjoyed being a teacher by picking up others' mistakes" (T2:19). From

her perspective, such student-student interaction could “create a better learning atmosphere than merely the teacher’s talk” and might also “create opportunities for students to self-evaluate their answers while they respond to others” (T2:19). However, it should be mentioned that students’ feedback in T2’s class remained restricted to whether their peers’ answers to the closed tasks were correct or incorrect, and the process involved in accomplishing the tasks was hardly attended to. Such student feedback might be a result of the closed tasks and questioning T2 adopted, in which correctness was the aim. It might also be the case that students did an imitation of how T2 provided feedback.

5.4 Understanding of assessment

After reporting on the teachers’ assessment activities, this section focuses on the teachers’ understandings of assessment. The interviews explore the teachers’ opinions regarding the assessment activities they conducted (section 5.4.1), the concept of formative assessment (section 5.4.2), and the influential factors that might have contributed to their understandings (section 5.4.3).

5.4.1 Different assessment activities

In the interviews, the two teachers were asked about their views of the assessment activities they employed. The first assessment activity was an oral task, the most often observed one in both teachers’ classes (as displayed in section 5.2.2.1 and section 5.3.2.1). According to T1, oral tasks served two purposes in her class. The first purpose T1 mentioned was improving students’ concentration. She argued that “students could be easily distracted in class”, so teachers needed to “raise questions to the distracted students and draw their attention back” (T1:5). T1 also added that “… I also want to check their understanding of the knowledge items” (T1:5). She reported that “I would invite a few students to see how they go; if they don’t understand, I will continue to explain” (T1:5). T2 agreed that she would “use oral tasks” to check students’ understanding when she “had enough time” (T2:5). However, she believed that oral tasks were “not an efficient approach”, as “some students hesitated too long without saying a word and wasted a lot of time” (T2:5). She also argued that it was often the case that when a few students responded to her questions, many other students became “distracted”, as “not being involved in classroom interaction was an excuse for them to not be examined” (T2:5).

The divergence between the two teachers' attitudes towards oral assessment activities appeared to stem from their perceptions of what they should achieve through oral tasks. T1 seemed to consider the primary purpose of oral tasks was tackling students' distraction problems. Checking students' understanding was also attended to, but from T1's description, enhancing students' focus seemed to come first. At the same time, T2 considered oral tasks time-consuming and regarded students' distraction as a result of such a method because it was difficult to ensure all students got involved in the tasks. She appeared to expect that one assessment activity could elicit all students' responses in a quick and organised way. What is more, she seemed to expect students to respond in a fast and correct way, as she regarded students' hesitance during their responses as unfavourable. The information underlay their hesitancy, which related to students' learning, did not seem to have been appreciated. Despite the distinct attitudes, the two teachers seemed to share one goal in their oral assessment, which was maintaining classroom discipline. T1 expected her students to stay focused in class, while T2 anticipated her students to perform unanimously according to her instructions so that all students remained concentrated. Their intentions, which I believe, were to help students overcome distraction or to ensure every student's involvement. However, the purposes of their activities were not primarily enhancing learning itself, and it was questionable whether such activities could bring students forward in their learning.

The second assessment activity was written task. T1 argued that written task, such as dictations and exercises (as displayed in section 5.2.2.2 and section 5.3.2.2), was an assessment method that could "assist teachers in quickly understanding how students have learnt, for example, after a lesson" (T1:6). According to T1, marking students' worksheets could allow her to identify "the problems students struggled with during the recent period" and "explain these problems in class" (T1:22). Likewise, T2 believed that written task was a practical assessment method that could elicit all students' responses and assess their performance altogether. From her perspective, it was unavoidable that "some students were incredibly shy and did not respond to any questions teachers raised" (T2:16). Therefore, T2 preferred asking students "to do the same thing at the same time, so everyone needs to prepare and perform" (T2:6), and teachers could "identify the problems from their responses" (T2:9).

Both teachers stressed that written tasks were manageable ways that could support teachers in attaining information about students' learning. T1 argued that written tasks could

efficiently assess students' learning at the end of a certain period, for example, checking students' memory of vocabulary after three vocabulary lessons. She seemed to perceive written tasks as a tool that could summarise students' performance regularly. At the same time, T2 highlighted how written tasks could involve more students and assist teachers in making judgments about students' performance. She seemed to prefer written tasks to oral tasks, as written tasks matched her expectation that all students could be assessed in a unified and simultaneous way. From the teachers' narratives, being manageable appeared to be an advantage of written tasks in their conceptions. Still, they did not mention how written tasks could assist students' learning.

The third assessment practice was student-assessed activity. T1 argued that she hardly employed student-assessed activities in class, which corresponded with the observational data. Her argument suggested that she was sceptical about the effects of such activities:

> Self- and peer assessment could be a fancy way of motivating students… but my students tend to assess others' performance by their relationship, not by objective criteria, and few can accurately assess themselves… For a few self-regulated students, self-assessment could remind them about how much they have learned, but for other students, the impact of such assessment can be negligible. (T1:6)

T1 appeared to believe that her students lacked the skills and knowledge in assessing others and themselves. She considered student-assessed activities as mere techniques that sparked students' motivation and denied the potential for these techniques to assist teachers' teaching, as "teachers do not assess students based on how they assess themselves; teachers use approaches such as dictations to assess students" (T1:23). From T1's perspective, students' objectivity in assessment and self-regulation skills determined their assessment capacity. However, she did not mention any attempt to promote students' assessment capacity and seemed to consider assessment as teachers' job, not students'. The reason why T1 highly valued objectivity might be associated with how Zhongkao tended to be 'objective tests', which included closed tasks such as multiple-choice and gap-filling. Reliability of the assessment tasks seemed to attract more attention from her, even though for classroom assessment, how effectively students' responses were interpreted and whether effective feedback could be provided matters more; reliability might be less central as a concern (Stobart, 2012). Such views of assessment have a clear behaviourist orientation, which will be discussed in Chapter 7.

T2 demonstrated a lack of confidence while she described her implementation of student-assessed activities, even though her students were noted to have assessed and provided feedback to their peers' responses with or without her instructions. T2 considered that "the classroom discipline issues" was her main concern while arranging student-assessed activities (T2:4), as in her previous attempts to organise group work activities, she found that many students were "merely chatting about random stuff instead of getting involved in learning" (T2:4). However, she recognised the positive influence of directing students to assess their peers with guiding questions or students assessing others spontaneously, as she commented:

> Students sometimes enjoy being a teacher by pointing out others' mistakes. If I correct all their mistakes, they are not interested; but if they are allowed to correct their classmates' mistakes, they are very keen to participate. (T2:19)

The switch of roles could be a key reason for students to participate actively in such classroom conversations. From T2's perspective, these conversations could be decent opportunities for students to "realise their problems" and "review what they learnt", and a "sense of getting involved" could also "positively influence students' learning motivation" (T2:19). However, T2's lack of confidence and concern for discipline hindered her further attempts to employ more student-centred assessment methods. Her lack of confidence might stem from her limited teaching experience and her reliance on senior teachers' instructions, and her concern for discipline might be influenced by the widespread emphasis on classroom discipline in her working environment.

5.4.2 Formative assessment

In the before-observation interview, the teachers were asked about their understanding of formative assessment and how formative assessment might influence their teaching and students' learning. T1 showed uncertainty about the definition of formative assessment by asking, "What do you mean by 'formative assessment'? Do you mean 'process assessment' (过程性评价)?" She reported having "heard about this term in lectures" and further commented that "the term 'process assessment' was more transparent" for her, as her interpretation of formative assessment was "assessing students during the process" (T1:8). This corresponded with other studies done in the Chinese tertiary English language education sector, where many teachers referred to formative assessment as 'process

assessment', which means "continuous and ongoing (assessment) as opposed to a one-off event at the end of a certain period", with "very little articulation of the procedures and purposes of formative assessment in relation to teaching and learning" (Chen et al., 2013, p. 840). T1 reported that she often conducted 'process assessment' by "encouraging students with positive comments", as encouragement could help "boost students' confidence" and "nurture their interest in English learning" (T1:8). She seemed to believe that formative assessment meant using encouragement to motivate students during their learning process. Students' current level, feedback on students' performance and learning process, or suggestions for improvement were not regarded as elements that featured formative assessment activities. Such a conception corresponded with T1's classroom conducts, in which assessment activities were largely teacher-centred and concentrated on making judgments of learning instead of promoting future learning. It also matched T1's feedback patterns, as confirmatory feedback was noted as the most widely implemented in her class, while feedback on students' future development was hardly noticed.

As for the influence of formative assessment, according to T1, formative assessment, or 'process assessment' as she called it, was a tool that helped teachers improve students' test results. As T1 argued, schools evaluated teachers' work by students' test results, so teachers "cared about students' scores" and would "adjust their teaching according to students' performance … for better test results (T1:8). T1 also argued that formative assessment had a positive impact on students' learning. She believed that "students concerned about scores and wished to be praised for the progress they made" (T1:8). If students received positive comments or results, they would "stay motivated to learn"; if they failed a test, they would "try harder next time to strive for a higher score" (T1:8). From T1's perspective, teachers' 'formative' actions, including obtaining information about students' learning, improving teaching, and encouraging students for their good performance, were for the purposes of achieving better test results, and test result itself could promote future learning. Such statements further demonstrated T1's summative-oriented assessment conception and her test-oriented teaching objectives.

T2, on the other hand, was unfamiliar with the term 'formative assessment', and I needed to explain the term to her before asking any further questions. Provided the explanation, T2 mentioned "regular dictations and quizzes" as her formative assessment activities, as these tasks could help her "regularly check every student's performance to understand whether

their learning is improving" (T2:5). In addition, she also considered praise as a type of formative assessment, as it was crucial for her to "identify every progress students made during the learning process" and use "every possible way to encourage them to move forward in their learning" (T2:6). From her statements, T2 appeared to consider regular mini-tests and encouragement representations of formative assessment activities. She seemed to understand formative assessment from a restricted perspective, in which frequent testing, teachers' corrections, and encouragement that contained little task-related information were viewed as formative (Carless, 2011; Hattie & Timperley, 2007). This matched T2's observed practices, as she implemented classroom assessment activities, particularly written tasks, as continuous snapshot takers of all students' learning processes and repeatedly adopted confirmation as a feedback approach. Assessment activities in T2's class performed a more summative function rather than a formative one, as the activities were organised primarily for making judgments instead of guiding future teaching and learning.

Regarding the influence of formative assessment, T2 focused more on how formative assessment could guide the feedback she provided to students. According to T2, different students should be assessed with different criteria: "those students who did well all along should try to avoid mistakes that should not be made, those average students are allowed to make some mistakes, and those who did not do well should be praised every time they achieve progress" (T2:7). From T2's perspective, encouraging students of different levels to make progress was more effective in promoting their learning than requiring all students to get all answers correct. Specifically, she argued that such practice could be particularly beneficial for those students who had lagged behind. As she commented, "although praise itself may not improve students' test scores directly, it could tell students that they are not good for nothing; if they study hard, they can always learn something" (T2:7). T2's attempt to customise criteria and feedback for students of different language levels and encourage them to make progress based on their current developmental level suggested that she recognised the learning-oriented function of assessment. However, from her narrative, it seemed that average students received less attention than the other student groups, as T4 merely suggested that they were "allowed to make mistakes" but did not put forward detailed plans for how their learning could be enhanced. The feedback T2 provided to students was also restricted to immediate error corrections at the task level and encouragement at the personal level, which might dilute the for-learning potential of the feedback and fail to bring about improved learning strategies (Hattie & Timperley, 2007).

5.4.3 Experiences related to assessment

The teachers described their experiences related to assessment in the interviews. They started with their past professional learning opportunities in assessment. T1 reported that although she had participated in numerous teacher education programmes before and throughout her career, most of the programmes focused on "promoting innovative pedagogic approaches, students' learning strategies, and introducing teaching and learning theories instead of assessment" (T1:9). While for T2, as a recent graduate who had limited academic or professional experience in English education, she reported that "classroom assessment has never been emphasised" either during her preparation for the teaching certificate or the professional education she received in School A (T2:7). It seemed that assessment had been a somewhat neglected topic in their previous education experience, which corresponded with studies on language teacher education in other contexts (Fulcher, 2020). Assessment, in this context, seemed to be considered separate from pedagogy and was not regarded as the responsibility of teachers. This corresponded with what was stated in the teacher guidebook, in which suggestions were mainly made for pedagogical procedures instead of assessment and feedback. A possible reason for this was that stakeholders in this context tended to associate assessment with external testing designed and administered by external examination boards. Such a phenomenon might explain the teachers' limited knowledge of assessment and less-developed capacity to implement formative assessment.

The teachers also discussed their present experiences with classroom assessment. Both teachers reported that their teaching materials, including textbooks and teacher guidebooks, provided few suggestions or examples of classroom assessment. T1 commented that the curriculum standards merely offered teachers "vague statements regarding utilising diverse assessment methods" (T1:13-14), and teachers needed to "try out different assessment approaches by themselves" (T1:14). T2 stated that she "did not notice" the assessment guidelines in the curriculum standards, and the teacher guidebook "only provided sample lesson plans, not assessment plans" (T2:11). Regardless of the external support, both teachers mentioned that they received peer support from other teachers. T1 reported that English teachers in School A gathered on a regular basis to discuss and share their "successful teaching and assessment experiences" in class (T1:7). T2 also stated that she benefited from the mentoring system School A established, in which "novice teachers were assigned experienced teachers as their mentors" and were encouraged to "observe their

mentors' teaching" and to "consult senior teachers about any questions they encountered" (T2:8). The teachers' descriptions showed that a teacher learning community had been created to facilitate their continuous professional development. However, it was questionable whether teachers could realise the negatives in traditional assessment methods and develop novel assessment activities to enable a more efficient and interactive language classroom without new assessment theories and knowledge being introduced to the community.

Besides their past and present experiences, the teachers were also interviewed regarding what role assessment played in their future actions. T1 emphasised the role summative tests played in the educational system and how these tests affected her future teaching and assessment. From her perspective, the Chinese educational system featured different types of summative tests, including unit tests, final tests, Zhongkao, and Gaokao, and it has been a custom that "teachers used tests to examine students' performance" and "relied on tests to make further decisions about teaching" (T1:15). While she acknowledged that "it took a long process to achieve an outcome", she found it "difficult to assess this process quantitatively" and further commented that "assessment (in our system) focused more on results" (T1:8). T2, on the other hand, stressed that students needed to perform well in Zhongkao to promote to public high schools. She said that under such a condition, "getting a better Zhongkao score" was more pressing for students than "acquiring a good command of English" (T2:13). It was, therefore, inevitable that "purposes of learning and teaching were more test-related" and "summative tests were more valued in school" (T2:12-13). Both teachers referred to summative tests, especially high-stake tests such as Zhongkao, as having an impact on their future practices. Their comments corresponded with their classroom assessment activities, wherein the constructs were selected based on their interpretations of what mattered in Zhongkao and their assessment conceptions, in which assessment performed a more summative role.

As a recurring theme in the dataset and a part of the assessment policies that exerted extensive impact, Zhongkao showed up as a key feature of the teachers' context. This directed the research attention to the policies framing the teachers' working environment. Several questions should be raised to understand the local assessment policies and construct a better understanding of how teachers reached their practices and conceptions under the framing of such policies: What were the school's requirements for teachers' classroom assessment? How did the Zhongkao design influence teachers' practices? Was there a

national guideline that informed teachers of what to do? The next section will address these questions.

5.5 Understanding of assessment policies

The chapter has explored the teachers' classroom assessment activities, their understandings of assessment activities and concepts, and how the teachers' assessment experiences impacted their assessment conducts. This section zooms out to discover how the teachers understood and interacted with the wider policy context and explore teachers' enactment of assessment policies at different levels and aspects, including the school policy (section 5.5.1), the Zhongkao policy (section 5.5.2), and the curriculum policy (section 5.5.3).

5.5.1 School A's policy regarding assessment

The two teachers were invited to share their experiences related to assessment policies from the school level and reflect on how such policies have influenced their work. Both teachers mentioned teacher evaluation as a critical school-based assessment policy. T1 reported that School A rated teachers as 'outstanding', 'pass', and, on rare occasions, 'failed' at the end of a semester "mainly based on students' test results in formal examinations, such as mid-term tests and final tests" (T1:11). From T1's perspective, being selected as 'outstanding teacher' was "an honour" and "a representation of self-actualisation" (T1:11). As T1 suggested, such an incentive measure could encourage teachers "to sharpen their teaching skills" as "they wished to receive positive remarks from students, parents, and colleagues" (T1:11). T2 stated that school leaders would "praise or criticise teachers at the semestrial meeting" according to their students' performance in formal examinations, which could be "a source of pressure" (T2:9). However, she also commented that "teachers' everyday teaching was hardly evaluated", so "the impact of such policy was limited" (T2:9).

School A appeared to place special attention on students' test results and has taken such results as the sole criteria for teacher evaluation. Instead of instructions, classroom activities, and feedback, scores were considered the indicator of the quality of teachers' work. The influence of the policy seemed to vary with each teacher. T1 showed greater enthusiasm for the 'outstanding teacher' title and considered it an aim worth striving for. T2, on the other hand, saw the policy as both a source of pressure and a restricted push for changes. It was difficult to conclude whether teacher evaluation might influence experienced and novice

teachers differently, as different people might interpret and respond to such policy in various ways. A conjecture might be that if a teacher found titles such as 'outstanding teacher' tempting, one might endeavour to meet the criteria underlying such a reward.

5.5.2 The influence of the Zhongkao reform

In the interviews, both teachers spent considerable time explaining the major changes in the current Zhongkao reform and how the new English Zhongkao was affecting their teaching and assessment. The teachers reported that the major changes in the English Zhongkao involve two aspects. The first change referred to the increase in the score of the listening-and-speaking test. The previous listening-and-speaking test accounted for 15% of the total score, and the proportion has been increased to 25% in the new test. The new test would take place through an online automated-scoring system, while the previous test was manually scored by markers, who were English teachers recruited across the city. The second change referred to the removal of vocabulary tasks and the subsequent increase in the score of reading comprehension tasks. According to the teachers, the vocabulary tasks, mostly comprised of multiple-choice tasks, were considered "incapable of reflecting students' language performance" and were therefore removed from the new English Zhongkao (T1:2). Instead, reading comprehension tasks increased to 40% of the total score, making the reading section account for the most significant share of the English Zhongkao.

The teachers discussed the influences of the above changes. As the teachers described, the most notable effect was the shift in teachers' teaching and assessment of speaking and listening. As T1 reported, before the Zhongkao reform, English teachers in School A "only started asking students to do listening-and-speaking mock tests in class from grade 9" (T1:12). However, due to the increased proportion of the listening-and-speaking test, "the teachers needed to start preparing from grade 7" (T1:12). According to T1, teachers mainly resorted to two approaches in "training students' listening-and-speaking skills" (T1:12). The first approach was assigning listening-and-speaking mock tests in class. The classroom observation showed that both T1 and T2 adopted listening-and-speaking mock tests from the workbook and required students to complete the test in class. However, the mock tests were treated more like listening exercises in both teachers' classes since the teachers asked their students to write down their answers on paper while they listened to the recordings and read their answers after they finished. A possible reason for such an arrangement was that treating

the mock tests as listening exercises was easier to administer in a large class, as having fifty students responding to the recordings simultaneously as they did in actual listening-and-speaking tests might seriously detriment classroom discipline. Another reason might be that the teachers considered students incapable of producing oral responses right after finishing the recordings. Writing the answers down might assist students in formulating their answers.

The second approach was students doing listening-and-speaking mock tests online after class. School A required all English teachers and students to use an online practice and preparation platform. Teachers could use the platform to assign homework and mock tests and monitor students' progress, and students could finish the assignments online. T1 argued that assigning a listening-and-speaking mock test on the platform was a "more straightforward way of assessing students' listening-and-speaking", as the online platform could "automatically rate students' performance, in the same way as the automated-scoring system used by the English Zhongkao" (T1:12). T2 also reported that she assigned online listening-and-speaking exercises "every week" to students as "regular practices" (T2:10). However, both teachers admitted that the feedback students received from the platform was limited to the scores for each section of the tests, and that they seldom provided feedback to students based on their performances in these tests. T1 argued that "checking on each students' answers" would be "an overwhelming workload" (T1:12), while T2 suggested that she spent most of her effort "motivating students to finish the mock tests", as "encouraging them to speak" was already an "enormous task" (T2:10). The online platform has enabled students to gain easier access to after-class listening-and-speaking practices, yet it could not provide teachers further assistance except for score reports. Even though it was much more convenient for teachers to respond to recorded audio instead of real-time responses in class, the workload involved remained realistic issues that hindered teachers' willingness to offer individual feedback.

Another effect of the Zhongkao change appeared in teachers' vocabulary and reading teaching. As both teachers argued, the removal of the vocabulary tasks in Zhongkao significantly influenced students and teachers in School A. Since the vocabulary tasks only examined the basic knowledge about words and phrases and could be easily accomplished through rote memorisation, students in School A "tended to perform well in such tasks" (T2:9). As T2 reported, "it might be difficult to improve our students' language abilities in a short period, but they always do well in vocabulary tasks" (T2:9). However, without the 'simple tasks' in Zhongkao, students in School A might "lose their only advantage" (T2:9),

and teachers needed to “reduce the vocabulary tasks assigned to students” (T1:2). As a result, “solid reading comprehension exercises and cloze”, as the teachers stated, were heavily emphasised in teaching and assessment (T1:2; T2:9). The observation showed that both teachers had spent considerable effort assigning and assessing reading exercises in class. However, it was also noted that their classroom assessment did not exclude basic vocabulary exercises and assessing students’ memorisation of vocabulary. Such classroom practices might result from the Zhongkao design and teachers’ language teaching beliefs, in which vocabulary persisted as a linguistic element that was examined in tests and required continuous attention.

Although the writing task in the English Zhongkao has not experienced dramatic changes in the current reform, the task remained influential on teachers’ practices. Firstly, the relevance of writing tasks to Zhongkao seemed to be a critical factor for teachers in deciding whether the writing section of a unit should be included in their teaching. It was reported in the previous sections that both teachers omitted the writing section of the unit. Reasons such as “task type will not be tested in exams” and “the task was too easy and not worth the time to teach in class” were reported (T1:informal interview; T2:informal interview). Secondly, the task design, in which Chinese instructions were adopted, and students were required to include the main ideas mentioned in the instructions in their writing, might also affect how teachers teach writing. T1 reported a ‘sentences translation’ (句群翻译) approach, which Grade 7 and 8 teachers generally adopted for teaching writing, as she suggested. The approach involved designing the writing tasks by providing a few Chinese sentences, which were usually the translation of key sentences in a sample writing, and requiring students to translate the Chinese sentences into English. According to T1, such an approach responded to “how the writing task design in Zhongkao” and aimed to assist students in “building a sense of how to write an article” (T1:18). However, as T1 admitted, ‘sentences translation’ approach led teachers to focus more on “teaching translation instead of teaching writing” (T1:18). Students were taught to produce grammatically and lexically correct sentences but were not instructed on how they should generate ideas and plan their writing. It was highly doubtful that such an approach could help students develop proper writing skills.

5.5.3 The influence of the Curriculum Standards

The influence of the latest English curriculum standards on the teachers' teaching and assessment was also investigated. T1 reported that although the school provided a copy of curriculum standards to every teacher, the actual document that guided teachers' work was the 'Guidelines for Middle School Graduation Tests in Shenzhen' (later referred to as the Zhongkao guidelines). Unlike the curriculum standards, which were national guidelines composed and published by the Ministry of Education, the Zhongkao guidelines were local guidelines edited and published by the Shenzhen Admission and Examination Office and the Shenzhen Institute of Education Sciences (2021), which involved performance descriptions for middle school graduates, Zhongkao design and sample test paper of different subjects, including English. According to T1, she mainly relied on the Zhongkao guidelines because the document was "a more tailored one for middle school students and teachers in Shenzhen" and provided "requirements for middle school graduates and a detailed introduction to the Zhongkao design" (T1:13). Even though the requirements in Zhongkao guidelines were, in fact, the same as the performance descriptions for middle school graduates in the curriculum standards, T1 still described the Zhongkao guidelines as "more useful" (T1:13). An inference could be made that the introduction to the Zhongkao design and the sample test paper included in the document, rather than the performance descriptions, were of more interest to T1.

T2 did not mention the Zhongkao guidelines as her guidance document. Instead, she considered Zhongkao itself the leading driver of teachers' work. As T2 commented, the curriculum standards were "too basic compared to the actual requirements of the Zhongkao" (T2:11). T2 shared her experience of attending a briefing about the grading standards of Zhongkao. She recalled that some of the writing samples displayed in the briefing aligned with the curriculum's requirements for middle school graduates' writing ability. However, according to the grading standards of Zhongkao, these writing samples received merely mediocre scores. To receive a top score, students needed to use phrases and sentence structures that would only be taught in high school. Consequently, T2 believed that if teachers simply followed whatever the curriculum standards prescribed, "it will be difficult for our students to receive a satisfactory score and have access to a high school" (T2:11). The issue T2 raised exemplified how the selection purpose of Zhongkao might contradict its qualification purpose. While the Zhongkao guidelines described the test as an achievement

test, which intended to reflect the English level students reached during the middle school years, the fact that not all middle school students could have access to public high schools made the selection purpose more critical. Performing better than others, rather than achieving what the curriculum prescribed, became the aim of teachers and students, and referring to Zhongkao for more targeted learning and practice turned out to be a 'reasonable' choice, even though other learning goals specified in the curriculum might be neglected.

5.6 Summary

This chapter outlines the two teachers' classroom assessment activities, their understandings of assessment, and their understanding and enactment of the assessment policies impacting their context. The data showed that the two teachers shared a lot in common in their classroom assessment activities. Both teachers tended to adopt oral assessment activities more in class, despite their distinct attitudes towards such assessment methods. They generally resorted to closed questions and tasks to examine whether students could produce an answer they anticipated and placed most of their attention on assessing linguistic knowledge, particularly vocabulary and grammar, and reading skills, which they considered crucial in test-taking. Language was taught as a set of discrete targets of linguistic elements, and communication was not the focus of teaching and learning. The teachers also adopted a more teacher-centred approach in conducting classroom assessment and expressed distrust and lack of confidence in implementing student-centred assessment activities. They mainly offered students restricted feedback in classroom assessment conversations, which involved confirmation and error correction at the task level and hardly addressed students' learning process.

The data demonstrated that while the two teachers designed their assessment activities, 'discipline issue' and 'being manageable' were their top concerns. Both teachers intended to use assessment activities to regulate students' behaviours and preferred assessment methods that could be easily implemented in large classes. Given the classroom context in which the teachers needed to take care of around fifty students, it is understandable that teachers favoured certain types of assessment that were more convenient, and it is justifiable that they regarded classroom discipline highly, as classroom disciplinary problems may detriment students' motivation and achievement (Arens et al., 2015). However, the other side of the paper is that tight control of classroom discipline could make the classroom learning

environment less collaborative, which might close the door for students to seek their peers' support. The data also showed that the teachers had limited knowledge regarding formative assessment. T1 adopted a restricted understanding of 'process assessment', which viewed formative assessment as continuous assessment of students' performance without attending to their learning, while T2 related formative assessment to frequent testing and praise. They tended to use exam results as the indication of students' language level and as the supporting evidence for their espoused teaching objectives, suggesting that their assessment conceptions remained largely test-driven and score-focused.

Further investigations into the teachers' experiences related to assessment and the policies they worked with provided insights into understanding their assessment activities and understandings. Both teachers reported that they received limited professional learning experience regarding classroom assessment and were given inadequate support from the school, the official textbook, and the teacher guidebook. From their narratives, it seemed that, in the context they worked, there was not a consensus that teachers' classroom assessment should be a significant part of teachers' teaching or that it was necessary for teachers to improve students learning through assessment. Tests remained a major factor applied by the school to evaluate teachers and adopted by teachers to review their own teaching, and Zhongkao was taken by the teachers as their future goals. The assessment guidelines in the curriculum standards were entirely ignored by the teachers. One encouraging phenomenon identified in the interviews was that the teachers were involved in the local teacher learning community and benefited from the support and experiences shared by their peers as teacher-as-learners. It was unsure whether such a community has improved teachers' practices, but it could be a good start for changes.

Chapter 6 Case study two: School B

Case study two adopts both data-driven inductive and theory-driven deductive approaches to analyse teachers' assessment practices and their understanding of assessment and assessment policies within the setting of School B. It starts with a profile of School B (section 6.1), presenting the relevant information to the study of the two teacher participants, T3 and T4, and School B, the context where the teachers worked. The teachers' comments on the students' language level and the sitting arrangements of the classrooms were also introduced. The following sections detail the assessment activities employed in the classrooms of T3 (section 6.2) and T4 (section 6.3). Section 6.4 documents the teachers' understandings of assessment, and section 6.5 reports on their understandings and enactment of the policies that impacted their assessment activities. Section 6.6 provides a summary of the findings of this case.

6.1 Profile

T3 was an experienced English teacher with over 20 years of teaching experience. As a senior staff member in School B, T3 took on several roles, including teaching and managerial roles. She was an English teacher of two classes, the head of the English teaching group of grade 8, and the head of the English department of School B. Having majored in English education in a teaching college, T3 received over five years of academic and professional education in English teaching and learning. Before becoming a full-time teacher, she had the opportunity to become an intern teacher and observe lessons in top schools in her home city. As T3 commented, the intern experience has nurtured her teaching belief that language should be taught in a meaningful context, and such a belief has been implemented throughout her practices in School B.

T4 was an early-career teacher who had only started her second year of teaching when the study was conducted. She obtained a master's degree in translation and interpretation studies from a renowned foreign language studies university in China and joined School B as a middle school English teacher after graduation. Despite many years of academic experience in learning English, T4 received limited academic or professional education in English teaching. Before graduation, her most relevant English teaching experience was part-time teaching jobs in tutoring institutions and obtaining a teaching certificate. Similar to School

A, School B also had a mentoring system, which assigned a mentor, usually an experienced teacher, to novice teachers so that they could learn from the experiences of the seniors. T3 happens to be T4's mentor, which added an interesting perspective in observing their activities and analysing their underlying beliefs.

School B was a public school that provided both primary school education and middle school education. It had an enrolment of around 1200 students in the middle school department. School B had 24 classes in the middle school department, eight classes in each grade. Each class has around 52 students. Similar to School A, each middle school English teacher in School B was in charge of two classes, that is, around 104 students in total. T3 and T4 commented on their students' level of English and some of the problems they encountered in their teaching. Both teachers reported a "polarisation" in their students' classroom and test performances (T3:1; T4:1). By 'polarisation', they meant there was a vast difference between the English level of different groups of students in the classes I visited. According to the teachers, both classes have "around ten students who are high-performing and responsive" in class and "around 15 students who perform poorly" in learning attitudes, classroom engagement and tests (T3:1; T4:1). This contrast with the reports from School A, as T1 suggested that "only two to three students" in one class performed well in classroom interactions and tests (T1:1). When asked how they would compare students' level of English with other schools in the district, both T3 and T4 reported that the English test performance of School B was not outstanding compared to other schools but was improving in recent years. As the head of the English department of School B, T3 commented that one of the reasons why the test performance of School B did not stand out was that the "teachers did not impose excessive work on students" due to the pressure of exams (T3:2). According to T3, over the years, teachers in School B were "learning advanced theories of language teaching and learning" to guide their practices and were trying "not to be entirely controlled by the exams" (T3:2).

The teachers also reported on how the classroom layout was arranged in the two classes, in which students were observed to sit in rows, facing the screen, blackboard, and the teacher's desk at the front. Both T3 and T4 were subject teachers (科任老师) of the two classes and did not oversee the student administration, so they were asked to inquire about the classroom layout with the class teachers of the two classes. The class teacher of T3's class arranged the seats based on two aspects: "students' height and characteristics" (T3:informal interview).

Taller students would usually be arranged to sit at the back, while chatty students would usually be arranged to sit with quiet students. The class teacher allowed students to choose their seats and whom they sat with "on the condition that they made a certain level of progress in their exams" (T3:informal interview). The seating arrangement was reallocated monthly.

The class teacher of T4's class also adopted "students' height" as the criterion for seating arrangement and "changed the seating plan every month" (T4:informal interview). Based on "a fairness principle", students who sat in the middle of the classroom would later be arranged to sit at the two sides of the class, and students who sat at the back would be given opportunities to sit at the front (T4:informal interview). To "maintain the discipline" and to "foster a better learning atmosphere" in class, students who could collaborate with and mutually support each other in their learning would be placed together, while students who enjoy having chats irrelevant to learning with each other in class would be separated (T4:informal interview). "Higher-performing students" and "students who made remarkable progress in exams" would be "given the opportunities to choose their own seats" (T4:informal interview).

From the teachers' report, it can be discerned that the class teachers considered discipline a crucial criterion that guided their arrangement of seats. They believed that discipline promoted a positive learning environment, and precedence should be given to discipline instead of students' preferences. This might be a result of the large size of their classes, which restricted their management choices regarding how much freedom they gave to students. However, it did not mean that students' wills were entirely ignored. In both classes, teachers offered conditions that students would be entitled to decide their seats if they performed well or made progress in tests. Test results have been used as an incentive to encourage students to work harder. This solution was, perhaps, considered effective in motivating students to learn. The class teacher of T4's class also shows an appreciation of the values of collaborative relationships between students in learning, as he took the rewarding sides of students' relationships into consideration. Although the class teachers retained the decisive roles in determining the classroom layout, they attempted to encourage students to take an active role in learning with various approaches.

6.2 T3

This section reports on T3's teaching and assessment in the classroom. Section 6.2.1 introduces T3's intended teaching objectives of the unit and reviews her actual teaching arrangements, which demonstrate the background of her assessment designs. Section 6.2.2 illustrates and analyses the observed assessment activities in T3's class.

6.2.1 Background of assessment design

The lessons observed in School B were based on unit 4 of the textbook, around the theme 'inventions'. T3 introduced her teaching objectives of the unit as threefold. The first one she mentioned was understanding the textbook passages about inventions. T3 believed that, from learning the passages, students should not only "practise their reading skills" but also "learn to describe an invention from different perspectives", for example, "what the invention was like, what changes the invention has brought to the world, and what the relationship between the invention and people was like" (T3:18). The second one was the unit's grammar points, including the irregular comparative and superlative forms of adjectives and phrases such as '(not) as … as'. T3 argued that students should learn "how to use these grammar points when they describe things around them" (T3:18), for example, their campus. The third one was learning "how to write about inventions" (T3:18). T3 stated that she intended to encourage students to write an article called 'My Invention' "based on the language knowledge of this unit" and "their life experience" (T3:18). From her statement, it seemed that T3 placed a strong emphasis on learning to use language in her teaching. She appeared to adopt a more communicative-oriented view of language learning, in which language was considered a tool for communication rather than merely linguistic knowledge (Halliday, 1985; Hymes, 1972).

As for the teaching arrangements of the unit, T3 adopted the textbook, the teacher guidebook, and the corresponding CD that provides downloadable slides and recordings as her lesson planning resources. The reason for T3 to adopt these official teaching materials was that "they shared a consistent voice regarding the goals and philosophy of teaching" (T3:3). From T3's perspective, teachers should plan their teaching according to the textbook structure and try their best to address all unit sections, including reading, writing, grammar, listening, speaking, more practices, and cultural corner, as all these sections were "internally connected" and "constructed a context" for students to understand and practise the target language

knowledge and skills (T3:3). She described her approach as a "contextual approach", in which "language should be learnt and taught in a context", and students could "connect with what they learn and construct knowledge within the learning process" (T3:6-7). According to T3, such a belief might contradict some teachers' practices, as they "tended to focus mainly on the reading, grammar, and writing part of each unit" because "the listening part in the textbook was considered by many teachers as too easy, while the speaking part has been given little attention all along" (T3:3).

The observation showed that, unlike the two teachers in School A, who omitted several sections of the unit, T3 had included most of the unit sections in her teaching, which largely corresponded with her assertion that teachers should address all sections of the textbook. She attributed her eight lessons to teaching and practising different language knowledge and skills in a relatively balanced way according to the textbook layout (one lesson for vocabulary, two for reading, two for grammar, one for writing, one for listening and speaking tasks and the project section, and one for exercises). It should be noted that T3 was the only teacher in this study who included the project section, an optional section at the end of the unit that encouraged inquiry-based learning, in her teaching. In the project lesson, T3 instructed students to review vocabulary of the unit and skimming skills with the article in this section, which was about an ancient Chinese invention. She then encouraged students to discuss how they could benefit from the invention and required them to do online research about other ancient Chinese inventions after school and write a booklet about an invention based on their research. With this lesson, T3 managed to help students review and practise several language knowledge items and skills, including reading, vocabulary, speaking, and writing. It could be seen as an example of T3's 'contextual approach' because she attempted to relate the target skills of the unit within a task and guide students' learning under a specific context. T3 displayed lesson aims adapted from the 'getting ready' section at the beginning of many of her lessons but omitted the 'self-assessment' section. According to T3, the 'getting ready' section could "inform students about what they are going to learnt", and the 'self-assessment' section should be assigned for students to "complete the form individually" to "save time for teaching other sections" (T3:9).

6.2.2 T3's assessment activities

Through classroom observations, T3's classroom assessment activities were recorded and reviewed. Among the eight lessons observed in T3's classroom, 90 assessment activities were identified and grouped into three types – oral assessment activities (section 6.2.2.1), written assessment activities (section 6.2.2.2), and student-assessed activities (section 6.2.2.3). The percentage of each type of activity was reported in Table 6-1.

Assessment activity	Times	Assessment activity	Times	Assessment activity	Times
TO-RC	21 (23.3%)	TW-RE	7 (7.8%)	PO-RC	3 (3.3%)
TO-V	19 (21.1%)	TW-GE	4 (4.4%)	PO-G	1 (1.1%)
TO-G	13 (14.4%)	TW-VE	2 (2.2%)	SW-D	4 (4.4%)
TO-S	9 (10%)	TW-LE	2 (2.2%)	PW-WE	1 (1.1%)
		TW-WE	2 (2.2%)	PW-VE	1 (1.1%)
				PW-RE	1 (1.1%)
Teacher oral total: 62 (68.9%)		Teacher written total: 17 (18.9%)		Student-assessed total: 11 (12.2%)	
					Total: 90

Table 6-1 T3's assessment activities

6.2.2.1 Oral assessment activities

The observational data demonstrated that T3 tended to assess students' oral responses, as 68.9% of the assessment activities were implemented to stimulate students' oral responses. They also showed that T3 adopted oral assessment more frequently than the two teachers in School A, since the identified oral assessment activities in her class were a third more than T1's and T2's observed oral assessment activities. Similar to teachers in School A, T3 often raised planned or improvised questions to individual students and the whole class. What made her lessons different was that these questions included both display questions, where there were fixed answers, and referential questions, which often begin with wh- questions (who, why, what) and elicit more than one possible answer (Walsh, 2011). Among these questions, reading comprehension, vocabulary, grammar and speaking were the major assessed constructs. Reading comprehension and vocabulary were assessed the most often. For reading comprehension, T3 often raised questions related to the details or the general ideas of passages from the textbook and workbook to check students' understanding and practise their reading skills. For example:

> T3 first explains the reading strategy 'skimming'. Then, she asks students several questions, such as "what is the title of the passage", "what is the introduction",

> "how many subtitles here", "what are they", "can you read the first sentence and the last sentence together" to guide students through the skimming process. (T3:L1)

> T3 asks, "when was compass invented?" Students scan the text in the workbook and answer, "in the Han Dynasty". … T3 then asks an open question, "who use the compass?" Some students say "people". T3 asks, " what kind of people?" Some students say "Chinese people" or "people from the West". T3 laughs and says, "maybe we can call them travellers or explorers", and writes the two words on blackboard. (T3:L7)

For vocabulary, T3 usually examined students' memory of the pronunciation and meaning of words and phrases. Besides raising the "what is the meaning of …" questions, T3 adopted alternative ways to draw out students' responses, particularly by describing the words in English and asking students to recall the words. For instance:

> T3 asks, "what does 'fur' mean?" Students answer in Chinese, and T3 confirms the answer. (T3:L4)

> T3 shows a sentence from the reading passage on the screen, "you can make a phone call with others anytime, anywhere", and asks students to think about the question, "what invention is described by the sentence?" T3 invites one student to read the statement. After the student finishes reading, T3 asks other students to answer the question. The students answer, "mobile phone", and T3 confirms the answer, "yes, mobile phone, good, good job." (T3:L2)

> T3 asks, "If you have a personal computer at home, put up your hands." Only four students raised their hands. Then, T3 asks, "If you have a laptop at home, put up your hands." More students raised their hands. So, T3 says, "So, more students have a laptop, right? So we can say, which is … something be liked by more people …" Student respond, "More popular!" T3 confirms, "more popular, yes", and write the phrase "more popular" on the blackboard. (T3:L6)

As for assessing grammar, T3 tended to use example sentences to check whether students understood the grammar points of the unit. For example:

> T3 guides students through a grammar exercise on the textbook. The exercise demonstrates pictures of a dishwasher and a microwave with information about their heights and widths. T3 picks another student and asks, "let's look at the sentence (The dishwasher is as ___ as the microwave), what do we put between 'as … as', high or higher"? Some students say "higher", some say "high", and the student says "high". T3 confirms his answer and further explains the grammar point. (T3:L5)

> T3 reads the sentence, "what is ___, the PC or the laptop?" She asks students to fill in the blank and says, "any answer is ok". One student says "bigger". T3 says, "Bigger, what other adjectives can you think of?" Other students say: "smaller, better, worse, more convenient, more inconvenient…" T3 says, "yes, we can fill in many comparative adjectives here, right?" (T3:L6)

T3 also assessed students' speaking by incorporating open questions within activities targeting assessing vocabulary, grammar, and reading. For instance:

> T3 asks, "Can you list any other important inventions in our daily life?" One student says, "Paper!" T3 says, "Yes, of course! Why is paper a great invention?" Another student says, "we can do homework on it." T3 says, "Good idea! We can also keep important information on it, right?" (T3:L1)

> T3 invites one student and asks, "How tall are you?" She intends to elicit sentences that ends with an adjectives (similar to the sentences in today's dictation). The student answers, "Maybe I am … one hundred and seventy centimetres tall." T3 repeats his answer and says, "Ok, good job." (T3:L5)

Feedback involved in T3's oral assessment tasks could be classified into four types. The first type, confirmatory feedback, was the most common type of teacher feedback in T3's class. As some of the above assessment moments revealed, when students responded to T3's questions with qualified answers, if the question was a closed question, T3 confirmed students' answers with a straightforward positive comment, such as 'yes' or 'good job'; if the question was an open question, T3 usually confirmed students' answers and resorted to further comments or supplement other possible answers. T3's confirmatory feedback to closed questions was comparable to that offered by T1 and T2. The simple response might result from the restricted format of the question, in which limited information could be gathered through students' answers. Her feedback to open questions, however, was more informative in reminding students of the language knowledge involved in the task and offering them alternative responses. Such approaches might have the potential to broaden students' linguistic repertoire and extending their thinking, as students were encouraged to convey their intended meaning in their own words, and language was not treated as arithmetic with sole correct answers (Swain, 1985).

The second type of feedback was sharing criteria with students. T3 often articulated the 'full-sentence criterion', which required students to answer questions with complete sentences, during oral assessment activities to elicit better responses from students. According to T3,

such a requirement "helps students better understand sentence structures" and could "further improve students' performance in writing tasks" (T3:23). An example is displayed below:

> T3 invites one student and asks, "How high is the dishwasher?" The student says, "three … no no no …" T3 says, "Give me a full sentence". The student says, "It is thirty-seven centimetres high". (T3:L6)

The third type of feedback was corrective feedback. When students' responses were not satisfactory enough, T3 pointed out their problems with different approaches. One strategy T3 usually adopted was asking additional questions, which could serve as hints that lead students to adjust their answers. Another strategy T3 used was providing explanations or corrections straight away so that students realised their mistakes and reached a proper answer. Here are some examples:

> T3 points to a picture of mobile phone and asks, "How does this invention help us in daily life?" One student answers, "telephone can help us keep in touch with others anytime and anywhere." T3 asks, "Are you sure? Can you take a telephone anywhere with you?" The student changes her answer, "mobile phone". (T3:L1)

> T3 reads the sentence on the screen, "the students who have a smart phone are ___ than those who have a laptop", and says "more, or …" The students say "less". T3 says "less?" and writes "fewer" on the blackboard, saying, "Student is a countable noun, right? So we should say 'fewer' students." T3 asks students to read after her: "fewer students", "more students". (T3:L6)

While the above three types of feedback mainly concentrated on how students performed in a specific task, the fourth type of feedback moved beyond the task level and associated more with how students should engage with feedback to better facilitate their learning. For example:

> "A good way of study is that when I give feedback to others, you also review your answer and take notes on my response. Don't just listen to what I say, or it will be a waste of your time." (T3:L4)

Students were suggested to actively engage with feedback information and reflect on their answers to enhance their capacity to create internal feedback. This type of instruction primarily aimed to foster students' ability to self-direct themselves towards their learning goals. It could be categorised as feedback at the self-regulation level, which could potentially bring about the development of effective learners (Hattie & Timperley, 2007).

6.2.2.2 Written assessment activities

The observational data showed that 18.9% of T3's assessment activities were teacher-assessed activities that aimed to elicit students' written responses. These activities included exercises targeting different language knowledge and skills, including reading, grammar, vocabulary, listening, and writing, and were used to assess students while they learnt. The written tasks took different formats, including gap-filling, multiple-choice questions, short-answer questions, and article writing, and assessed students' reading, listening, writing, and understanding of grammar and vocabulary individually or integratedly. The following excerpts are examples of different written exercises assigned in T3's classroom.

Task 1
Read the article again and answer the questions below in complete sentences.
1. Which of the three things was invented first?
_____

Task 2
Listen to the recording again and circle the correct answer.
1. The first invention can ____.
A. make your shoes more beautiful B. keep the dust off your shoes
C. keep the rain off your feet

Task 3
Complete the conversation below with the words from the box. Change their forms if necessary.

anytime century comfortable daytime passenger

Maggie: People developed the first cars in the 1880s, and they started using cars a lot in the early 20th (1) _____.
Bill: I'm sure the people were very happy because cars are so fast and (2) _____.

Task 4
Write a short article about an ancient Chinese invention. Use the outline and the example below to help you.

These written tasks were mainly exercises from the textbook, workbook, and teacher-made worksheets. They were normally designed in relation to the topic of the unit (in this case, invention), the vocabulary and grammar taught in this unit, and addressed various language skills, such as reading, listening, and writing. According to the observation, T3 assigned

different written tasks as in-class exercises or after-class homework based on the focus of her lessons. In-class exercises were completed by students during class time, and T3 would require them to share their answers and provide feedback accordingly. After-class homework was collected and marked by T3, who commented on students' performance and explained the tasks to address the common mistakes they made. The interactional patterns involved during the exercise explanation were similar to those observed during the oral assessment tasks. For example:

> T3 gives students three minutes to complete the short-answer questions in 'More Practice' by underlining the corresponding text for each question and writing down the answers. After students finish, T3 invites students to share their answers and confirms their answers by sharing sample answers on screen. (T3:L8)

Feedback involved in T3's written assessment activities resembled that incorporated in her oral assessment activities, particularly for exercises with closed formats and fixed answer keys. For instance, confirmatory feedback was supplied if students responded with the correct answers, and corrective feedback with explanations was provided when students made mistakes. For written tasks with a more open format, such as writing a booklet or an article, T3 shared writing samples with students to explain what made good writing and what common mistakes could be made. The samples could be an extract from the teaching materials or a piece of writing from one of the students. For example:

> T3 shows a sample writing about invention on the screen, with the structure and the important phrases of the article highlighted. Then, T3 explains the phrases and the structure and suggests that students could learn from the sample writing. (T3:L6)

> T3 shares a booklet written by one of the students. T3 points out the well-written sentences of the booklet and the language mistakes made in the writing. (T3:L8)

Besides feedback on specific tasks, T3 also provided feedback on what she called the students' "learning habits (学习习惯)" (T3:8). I took the term 'learning habits' to be the proper learning strategies that students could adopt to enable ongoing effective learning, for example, doing regular previews and reviews, taking notes, and paying attention in class. Such feedback usually encourages students to develop these strategies to better facilitate their learning, support teacher's teaching, and improve their capacity for self-regulation. For instance:

> "Finish your homework and hand in your homework in time; that's how I know where your problems are." (T3:L2)

> "Remember to review and revise the mistakes I circled out in your homework, so you know your problems, and try to avoid those problems the next time." (T3:L3)

6.2.2.3 Student-assessed activities

Besides teacher-assessed activities, twelve student-assessed activities were noted in T3's class, which was a much higher figure compared with the number of student-assessed activities implemented by the two teachers in School A. Among these activities, four involved T3 instructing students to self-check their dictation against the answer keys. Students were required to mark their worksheets, revise their mistakes, copy the right answer a few times to enhance their memory, and hand in the revised worksheets for T3 to examine. They were also encouraged to exchange their worksheets with a classmate and mark each other's work. Such activities were more than self-marking because T3 offered detailed instructions on how students should deal with and learn from their mistakes. T3 has also collected information by gathering and examining the revised worksheets, which might provide evidence of students' learning and support her future lesson planning.

The other student-assessed activities in T3's class mainly involved students discussing with their partners about questions related to specific tasks or assessing others' work with given criteria. For instance:

> T3 asks students to discuss in pairs about reading comprehension questions in section d2 in the textbook. She walks around, listens to students' discussion, and later invites student pairs to share their answers. She confirms their answers by showing answers on the screen and asks other students to check their answers. (T3:L1)

> T3 asks students to assess their partners' writing according to an evaluation form on the screen. The form includes four criteria: "Is there any misspelled words? Is there any incomplete sentence or sentence that doesn't make sense? Does the writer use comparative or superlative adjectives or 'as…as'? Does the writer tell us what the invention is special about, how it works, and what it looks like?" Two sample comments were also provided: "I like this writing because …", and "This writing can be improved by …" (T3:L6)

T3 arranged discussion activities and peer-check activities to enable conversations and collaborations among students about the target language knowledge and skills. Particularly,

in the writing assessment activity mentioned above, she followed the teacher guidebook's suggestions on instructing students to "revise their classmates' writing in pairs" (Zhang & Shu, 2013, p. 40). Such activities created spaces for students to comment on others' performances, realise their problems, understand the criteria, and learn from one another. One problem in the design of these activities, however, might be that some of the tasks involved in students' discussions, especially those related to reading comprehension, were mostly closed tasks with fixed answers. The discussions around these tasks could only take place within a narrow topic and with limited lexical choices. Therefore, the feedback that emerged between students was restricted to either confirmatory or corrective feedback at the task level and was difficult to extend to the learning process or self-regulation level. However, it should also be noted that students' language level can be a factor that restricts teachers' designs of assessment formats. For students involved in this study, who were mostly at the elementary level (A2 level in CEFR), the teacher might only resort to limited assessment choices so that students were not given tasks that exceeded their current capabilities (Vygotsky, 1978).

6.3 T4

This section details T4's teaching and assessment in the classroom. Section 6.3.1 demonstrates T4's stated teaching objectives of the unit and her actual teaching arrangements, which introduce the background of her assessment designs. Section 6.3.2 reports on T4's classroom assessment activities.

6.3.1 Background of assessment design

T4 mentioned three learning objectives that she wished to achieve through teaching the unit. The first objective was "learning the expressions related to inventions" and "how to use these expressions to describe particular inventions" (T4:16). The second objective was "furthering students' understanding of the grammar points of this unit, comparatives and superlatives forms" (T4:16). The third objective was "using the language knowledge learnt in listening and writing, for example, writing an article about their own inventions" (T4:16). T4's intended objectives bore a similarity with T3's. Both teachers emphasised the mastery of language knowledge, such as expressions and grammar, and gave special importance to applying the language knowledge in different language skills, which displayed a meaning-

focused and usage-based manner (Liu & Wang, 2020). Such a sharing understanding might result from T3 and T4's relationship as mentor/student and head/member of the teaching group. During the observation, T3 held regular meetings among the English teaching group to discuss their teaching arrangement, and T4 observed T3's lessons regularly and used T3's lesson design for reference. T4 might have adopted T3's teaching principles through these meetings and observations and integrated them into her espoused beliefs.

Similar to T3, T4 also depended on the textbook and other related supporting resources to plan her lessons. Specifically, she mainly relied on "the slides from the corresponding CD" and "the slides designed and shared by other experienced teachers" to plan her lessons, as "the slides' design matches the textbook's design" (T4:2). However, unlike T3, T4 considered the teacher guidebook a less helpful resource for her teaching. While she admitted that the teacher guidebook provided teachers with clear guidance on what knowledge should be taught and emphasised, the classroom activity design provided, as she believed, could hardly be adopted directly in her class. From T4's perspective, some activities in the guidebook were "designed for high-performing students" and were "too complicated for her students", while other activities that fit the students' level were too dull and "could not attract their attention" (T4:2). As a result, T4 suggested that she preferred "designing her classroom activities depending on the lesson's content" and her "students' interest" instead of the guidebook's suggestions (T4:2).

As for the teaching of a unit, T4 argued that she would not strictly follow the textbook's design, as she would "plan the unit's teaching as a whole" and might "omit some sections to fit the schedule" (T4:2). For example, the reading section and the 'more practice' section might be taught as a whole, as both sections focused on reading; the 'cultural corner' and the project section might be skipped or only briefly explained if she considered them "not as important" as other sections (T4:3). The observation showed that, unlike T3 who addressed almost all textbook sections, T4 excluded several sections from the teaching plan. Among the nine observed lessons, T4 allotted one lesson for teaching vocabulary, three for grammar, two for reading passages, one for writing, and two for exercises. The project section was ignored because "the time was limited" (T2:5); the speaking section was taught in a morning reading lesson, in which students were required to read the sample conversations in the textbook; the 'getting ready' and 'self-assessment sections were left out, as T4 found listing aims and letting students tick the boxes themselves "not useful" (T2:6). Although T4 shared

comparable reported beliefs about learning objectives with T3, their actual teaching arrangements appeared to diverge. T4 tended to place more attention on teaching grammar, reading, and explaining exercises in her teaching, as reported in the following section. Her actual practices showed that despite her reported usage-based beliefs, she tended to focus on the forms of language and training discrete skills with exercises, a practice strongly influenced by the behavioural psychology and structural linguistics (Brown & Abeywickrama, 2019).

6.3.2 T4's assessment activities

T4's classroom assessment activities were identified and logged through the classroom observations. Among the nine lessons observed in T4's classroom, 112 assessment activities were found and sorted into three types – oral assessment activities (section 6.3.2.1), written assessment activities (section 6.3.2.2), and student-assessed activities (section 6.3.2.3). The percentage of each type of activity was calculated and reported in Table 6-2.

Assessment activity	Times	Assessment activity	Times	Assessment activity	Times
TO-V	23 (20.5%)	TW-GE	10 (8.9%)	PO-T	4 (3.6%)
TO-T	22 (19.6%)	TW-RE	3 (2.7%)	SW-GE	3 (2.7%)
TO-RC	16 (14.3%)	TW-LE	3 (2.7%)	PO-V	2 (1.8%)
TO-G	10 (8.9%)	TW-D	2 (1.8%)	PW-GE	1 (0.9%)
TO-S	5 (4.5%)	TW-WE	1 (0.9%)		
TO-RA	5 (4.5%)				
TO-R	2 (1.8%)				
TO total: 83 (74.1%)		TW total: 19 (17%)		student-assessed total: 10 (8.9%)	
					Total: 112

Table 6-2 T4's assessment activities

6.3.2.1 Oral assessment activities

The observation showed that, among the four teachers in this study, T4 adopted oral assessment tasks the most frequently in class (83 times in one unit of teaching, 74.1% of her total identified assessment activities). Similar to other teachers, T4 generally raised questions to either individual students or the whole class to elicit students' oral responses. She also tended to raise closed questions based on what the lesson was about to assess students' understanding of the lesson content but also adopted open questions occasionally to elicit students' thoughts. Table 6-2 shows that vocabulary, translation, reading comprehension, and

grammar were the main focuses of T4's questions. For vocabulary, T4 examined students' memory of the meaning or spelling of the words they encountered. For example:

> T4 explains the word 'practical', saying, "we've already learnt the verb and noun, remember?" One student says, "practice". T4 says, "the verb is p-r-a-c…" She and students say together "…t-i-s-e". T4 then asks, "how about the noun?" She and students say together "p-r-a-c-t-i-c-e". (T4:L1)

> T4 asks, "what's the meaning of funny?" Students answer in Chinese. T4 writes notes on the blackboard "fun – funny" (words learnt in the previous unit) and explains other related phrases (have fun, funny story). (T4:L2)

T4 also required students to translate both Chinese and English sentences, which she made up according to the textbook content, to assess whether students had understood the target language knowledge. For instance:

> T4 says a Chinese sentences "它可以让你家的地板没有灰尘" and invites a student to translate with the phrase "keep … off" they learnt today. The student answers, "it can keep the dust off your floor." T4 says, "Ok, very good!" (T4:L3)

> T4 asks a student to translate the sentence "how was the ball point pen invented" into Chinese. The student answers in Chinese. T4 then asks, "which paragraph is about this?" The student answers, "the second paragraph". T4 confirms the answer, "Ok, sit down please." (T4:L4)

Students' understanding of passages in the textbook and workbook was often attended to. Questions were raised by T4 regarding the key information in the passages to practise students' reading skills. For instance:

> T4 plays the recording of the second paragraph and asks questions regarding this paragraph: "Who invented the telephone? When was the telephone invented? What could people do after this invention?" Students answer with details in the paragraph. (T4:L1)

> T4 asks, "What has light been liked after the invention of the light bulb?" She invites a student to answer. The student reads from the text, "people can do as many things in the evenings as they can in the daytime." T4 confirms the answer, "yes, very good!" (T4:L2)

Grammar received close attention in oral assessment. T4 tended to assess students' understanding of grammar knowledge through questioning. For example:

> T4 asks one student, “what’s the difference between a few and a little?” The student says: “肯定否定 (affirmative and negative).” T4 says: “肯定 (affirmative)? Both are 肯定(affirmative).” The student then says: “可数不可数 (countable and uncountable).” T4 asks: “哪个修饰可数(which one is for countable)?” The student says: “a few.” T4 then asks: “那哪个修饰不可数呢 (which one is for uncountable then)?” The student says: “a little.” T4 confirms the answer, “that’s right.” (T4:L7)

Besides the above themes, T4 also examined their memorisation of selected texts from the textbook through recitation, assessed students’ speaking through open questions, and checked their pronunciation through reading aloud activities. For instance:

> T4 shows the Chinese translation of selected texts from textbook (learnt in the last lesson) and asks students to recite the sentences with the support of Chinese translation. (T4:L1)

> T4 asks one student, “which do you prefer, ballpoint pen or fountain pen?” The student answers, “I like ball-point pen.” T4 says, “I prefer…” The student says, “I prefer like …” T4 explains the meaning and usage of “I prefer sth”, and the student revises her answer, “I prefer the ballpoint pen”. T4 asks, “why?” The student answers, “It is more convenient.” T4 says, “It is more convenient, yes, good. Any other opinions?” She invites another student. The student says: “I prefer the ballpoint pen, too.” T4 says, “Yes? Why?” The student says: “because I use this invention every day.” T4 says, “Oh! You use this invention every day. Yes, very good.” (T4:L4)

> T4 asks students to read the example conversations on the textbook and walks around to check individual students’ pronunciation. (T4:L6)

The feedback T4 implemented was similar to that of T2, which mainly involved three types of feedback: confirmatory feedback, disconfirmation with hints or follow-up questions, and sharing the ‘full-sentence criterion’. The first type, confirmatory feedback, was the most frequently observed one in T4’s class. When students’ answers met her expectation, T4 normally responded to students with “yes, very good”, “that’s right”, “Ok, sit down please”. Such responses were commonly seen in all four teachers’ classes, particularly when students displayed that they knew the specific answers to the teachers’ closed questions. It was usually the case that T4 did not offer any further suggestions after she confirmed students’ answers at the task level. This might result from the closed questions adopted, which offered limited space for students to demonstrate their skills further.

The second type of feedback, disconfirmation with hints or follow-up questions, was also applied when there were problems in students' responses. Unlike T2, who would use questions such as 'are you sure' to call students' attention to the problems, T4 would usually point out students' mistakes directly by repeating their answers with a rising intonation, indicating that the answers were 'wrong'. Sometimes, the disconfirmation feedback was provided along with hints suggesting which aspect of language knowledge students should pay more attention to or with follow-up questions that examined whether students realised their mistakes and understood the language knowledge assessed through the tasks. Some examples of this type of feedback are given below:

> T4 asks one student to fill in the blank "one of the greatest __ (invention)". The student says "invention". T4 says, "invention?" The student quickly changed his answer to "inventions". T4 asks, "Why?" The student explains the grammar point. (T4:L2)

> T4 asks one student to complete the sentence, "Bell ____ (invent) the telephone." The student answers, "invent". T4 asks, "Invent? He did it before." The student then corrects his answer, saying "invented". (T4:L3)

Along with the hints and questions, such feedback offered additional opportunities for students to revise their answers and review related language knowledge. However, it should be noted that the straightforward disconfirmation T4 adopted denied the opportunities for students to reflect on their performance, which might lead to students developing a trial-and-error strategy, in which attention was given to the immediate tasks instead of building up strategies that develop stronger associations of linguistic patterns and spending more cognitive efforts in generalising the related utterances. Suggestions aiming at the language patterns involved in these tasks and the processing of these patterns might be more helpful in the long term.

The third type of feedback was sharing the 'full-sentence criterion'. Similar to other teachers in the study, T4 often asked students to answer her questions in a full sentence. According to T4, one of the common mistakes her students made in writing and speaking was that "the sentences they made had no subject or predicate" (T4:19). T4 believed that asking students to practise answering in full sentences could "help them understand the importance of sentence structures", which might "further improve their writing and speaking" (T4:19). Through sharing this criterion, students were given clear goals in terms of what their responses should be like. An example is presented below:

T4 asks one student, “what’s the top speed of the car?” The student says, “three hundred kilometres per hour”. T4 says, “You should use a full sentence. The top speed of the car …” The student then says, “The top speed of the car is three hundred kilometres per hour.” T4 confirms the answer. (T4:L8)

6.3.2.2 Written assessment activities

The observational data showed that written tasks comprised 17% of T4’s assessment activities. These tasks included dictations, which assessed students’ memory of words and phrases of the unit, and exercises selected from the textbook, workbook, and teacher-made worksheets, which involved gap-filling tasks, multiple-choice questions, short-answer questions, and writing tasks and examined students’ understanding of grammar knowledge and language skills, such as reading, listening and writing. The following excerpts are samples of written exercises adopted in T4’s classroom.

Task 1
Look at the pictures below and complete the following sentences.
1. The Hotwave and the Micro-King __*are not as expensive as*__ (expensive) the Speedcook and the Superoven.
2. The Speedcook _____(wide) the Supercook.

Task 2
Read the article on page 51 and complete the table below.

	The wheel	The telephone	The light bulb
Inventor	Don’t know	_____	_____
Life after the invention	Travelling became _____ _____	People can _____ _____	People can _____ as they can in the daytime

Task 3
Listen to advertisements for four funny inventions and complete their names. Write one word in each blank.
1. _____ For Shoes.
2. Cleaner Cat _____
3. _____ Telephone
4. _____ Glasses

Task 4
Write a short article about your invention. Use the example below to help you.

Similar to the other teachers in the study, T4 assigned written tasks as either in-class exercises or after-class homework. In-class exercises were finished by students during class time, while after-class homework was assigned for students to complete after class and later collected and marked by T4. T4 would explain the language knowledge involved in these tasks and comment on students' performance according to their mistakes. The conversational patterns observed during written assessment activities resembled the dialogues that occurred during oral assessment activities. For example:

> T4 asks students to check their answers on the workbook together. She asks, "Why should we use comparative form here?" Students answer together, "There is a 'than' in the sentence." T4 asks students to mark the word "than". She then asks one student to explain, "Why should we use 'few' in this sentence?" The student answers, "修饰可数名词(it's before a countable noun.)" (T4:L7)

The feedback types T4 implemented in written assessment activities were similar to those applied in her oral assessment activities. Confirmatory feedback was provided when students came up with accurate answers; disconfirmation with hints or additional questions was generally delivered if students made mistakes in their exercises, and explanations of the related language knowledge were offered; the full-sentence criterion was highlighted when students responded to short-answer questions. For writing tasks, T4 also shared sample writing and instructed students on the necessary writing steps before asking students to write their own articles. She considered that it was "difficult" for her students to "write an article without any support", and "examples and instructions should be provided to them in advance" (T4:21). Such feedback could potentially reduce the gap between students' performance and the desired goal attainment:

> T4 shares a sample writing on the screen, with keywords, phrases and mistakes highlighted. She reminds students about the format of the notice task and the steps involved in completing a writing task. (T4:L5)

6.3.2.3 Student-assessed activities

Besides teacher-assessed activities, ten student-assessed activities were observed in T4's class, which consisted of 8.9% of the observed assessment activities. Among these activities, four involved T4 requiring students to check their written answers in exercises individually or in pairs according to the answer keys given. After students finished marking, T4 explained the written tasks straight away without checking students' performance or providing

individual feedback. While T4 described such activities as “self- and peer assessment” (T4:5), the activities could only be classified as ‘self- and peer marking’, as little information was gathered by T4 to inform her instructions.

The other six student-assessed moments took place during several oral assessment activities, which examined students’ translation skills and vocabulary. Among these moments, five involved students assessing or supporting their peer when a student had trouble answering questions without specific instructions from T4, while one involved T4 encouraging students to assess their peer’s answer. For instance:

> T4 asks one student to translate a Chinese sentence with the phrase “one of”. The student answers, “He is one of the highest boy …” T4 says “tallest”. The student quickly changes the answer to, “tallest – tallest boy in our class”. T4 says, “boy?” Other students say “boys”. T4 asks, “What should be added – this is a noun, what form should be used? Ok, try again.” The student then answers, “He is one of the tallest boys in our class.” (T4:L3)

> T4 asks one student to translate a sentence with “not as … as”. The student says, “I run not as fast as he run.” T4 asks other students, “Do you agree with him? Which auxiliary verb should we use here?” One student says, “I doesn’t …” T4 says, “I doesn’t?” Another student says, “I don’t”. T4 confirms, “I don’t. I don’t run as fast as he can.” (T4:L6)

Among these activities, students acted as error detectors with or without T4’s guidance, which was comparable to some of the student-assessed moments in T2’s class. According to T4, such a phenomenon was commonly seen in her class, as students, “especially higher-performing students, enjoyed identifying their peers’ mistakes and shared their answers” (T4:20). However, from the observation, it seemed that T4 rarely took advantage of students’ enthusiasm in assessing others to facilitate more dynamic classroom interactions, in which students were included intentionally as the assessors. Raising more questions during classroom assessment activities, such as “do you agree with him/her”, might enhance students’ engagement. The closed task involved in these student-assessed moments might have also restricted the content and format of students’ feedback, as the correctness of their peers’ answers was their sole focus. Arranging more open discussions among students might enable students’ feedback beyond error corrections.

6.4 Understanding of assessment

After discussing T3's and T4's assessment activities, this section focuses on the teachers' understanding of assessment. The interviews investigate the teachers' opinions regarding the different assessment activities they conducted (section 6.4.1), formative assessment (section 6.4.2), and the influencing factors that might have impacted their understanding (section 6.4.3).

6.4.1 Different assessment activities

In the interviews, both teachers were invited to share their views about the assessment activities they employed. The first assessment activity was oral task, the most frequently observed one in their classes. T3 reported that she mainly used oral tasks to "encourage students to speak up" and check "whether students understood what has been taught" (T3:4). From T3's perspective, teachers should offer "both encouragement and prompts to students" so that those who did not respond properly could make progress (T3:4). However, she also admitted that not all students could receive guidance through oral tasks, as teachers could "only reach limited students through questionings" (T3:4). T4 considered oral tasks an essential assessment approach in class and explained three purposes that drove her to use such an approach. The first one was improving students' focus. According to T4, raising questions frequently about the lesson content could "keep students' from becoming distracted" (T4:4). The second one was increasing classroom interaction. T4 argued that oral tasks could "make classroom environment more active" and "create more teacher-student conversations in class" (T4:4). The third one was emphasising the important knowledge. T4 argued that by raising questions constantly on particular language knowledge, students might "realise the importance of the knowledge and memorise it by heart" (T4:4).

The teachers' reports demonstrated their distinct perceptions of how they viewed the functions of oral tasks and what they could accomplish via oral tasks. T3 believed that oral tasks aimed to enhance students' confidence and assist their learning through prompts. Such a belief corresponded with how T3 took a more student-centred position and provided positive and supportive feedback to students, as she was concerned about how students' motivation could be strengthened and how their responses could be improved. On the other hand, T4 seemed to adopt a more teacher-centred belief regarding the use of oral tasks. The

purposes she suggested were more associated with how she wanted students to perform, for example, staying focused, responding to her questions, and paying more attention to specific language knowledge. However, whether students' learning could be enhanced through the oral tasks and the subsequent feedback was not included in her espoused belief. It is worth noting that T3's concern about oral tasks was similar to that of T2, who considered oral tasks not efficient in supervising every student. The size of their classes appeared to make assessing all students together with oral tasks impossible, which they considered a problem. T4 also shared the belief about how oral tasks might tackle students' distraction problems with T1. The tasks served as both assessment and managerial solutions to classroom discipline issues in the two teachers' classes.

The second assessment activity was written task. Both teachers confirmed that they often used written tasks in class. T3 argued that written tasks could "help teachers discover and deal with the problems in students' learning in time" (T3:4). She also believed that written tasks could "boost students' confidence" since they were usually composed of relatively "easy and straightforward tasks" (T3:4). T4 reported that she assigned written tasks in class because they were "efficient ways to assess" students' attainment, as "students could finish a task in a short time", and teachers could "obtain a general overview of students' performance by collecting the worksheets" and "offer corresponding feedback soon afterwards" (T4:4). She also suggested that written tasks "set up small goals" that could be easily achieved and thus provided students with "a sense of achievement" (T4:4). Both teachers mentioned how written tasks could assist teachers in obtaining information about students' performance in a timely way. It seemed that the teachers considered written tasks an assessment tool that could be easily managed, which corresponded with T1's and T2's views, in which being manageable was considered an advantage of written tasks. The teachers also stated how written tasks might increase students' self-confidence. They appeared to believe that improving confidence could benefit students' learning.

The third assessment activity was student-assessed activity. Both T3 and T4 reported implementing student-centred activities to assess students' performance. However, their opinions towards such activities differed. T3 believed that student-assessed activity, particularly having students self- or peer-mark their written tasks and having students discuss in pairs, was beneficial for their learning, as "they could receive instant feedback", "learn from each other through conversations" and "learn to take responsibility for their study"

(T3:5). However, she also mentioned that if she had enough time, she would prefer checking students' work by herself so that she could "identify the problems" and "make subsequent changes in teaching" herself (T3:5). On the other hand, T4 remained sceptical about the impact of self-marking but highly regarded pair work. She argued that "some students might not be entirely honest with the scoring and could lose the opportunity to learn from their mistakes" and mentioned that she generally assigned the marking tasks to individual students when she "could not finish scoring after class" and preferred "marking by myself" (T4:5). Instead, she suggested that having students mark in pairs could "usually discover more problems", "engage them more", and "enhance their motivation" (T4:5).

Despite their different views regarding self-marking, the teachers appeared to demonstrate more positive attitudes towards student-assessed activity than those of T1 and T2 (the difference between their attitudes will be further discussed in Chapter 7). This corresponded with the observational data, as more student-assessed activities were identified in T3's and T4's classes. The discrepancy between the two teachers' attitudes towards self-marking might result from their different approaches to arranging self-marking tasks. T3 supplied detailed guidelines on what students should do during self-marking, particularly how they should deal with their mistakes, while T4 did not set specific requirements concerning what students should do with their mistakes. The former approach intended to foster students' abilities to self-monitor their learning and correct their mistakes, while the latter provided little useful feedback to students and might have limited influence on facilitating learning improvement. The corresponding effects of the approaches and whether the teachers realised the importance of nurturing students' self-regulatory proficiency might have impacted their views of the assessment approach. Another point worth noting was that despite the fact that the teachers recognised the significance of arranging student-centred assessment activities, they described such activities as a substitute for teacher-assessed activities. They seemed to prefer gaining information about students' learning themselves and taking control of the teaching and learning processes.

6.4.2 Formative assessment

The teachers were asked about their interpretation of formative assessment and how formative assessment might affect their teaching and students' learning in the before-observation interview. Similar to T1, the first reaction of T3 to the term 'formative

assessment' was "the 'formative assessment' you talk about … do you mean 'process assessment' (过程性评价)? Assessing the process of students' learning?" (T3:7) She reported to have "learnt about this term in a teaching training project" and stated that 'process assessment' was vital in teaching and learning, as a teacher should "use the assessment process to help students form better learning habits", which would "further lead to success in students' learning" and "lessen teachers burden" (T3:8). T3's reaction was somewhat similar to T1's, as they both considered 'process assessment' as an alternative term that described 'formative assessment', despite their distinct interpretations. While T1 considered formative assessment as positive comments by the teacher, T3 perceived the term as helping students develop appropriate learning strategies. Such a belief contradicted Chen et al.'s (2013) report on Chinese teachers' interpretations of 'process assessment', in which college English teachers viewed process assessment as ongoing judgements on students' performance. T3's interpretation of formative assessment, or 'process assessment, echoed her practices, as her feedback both concerned students' performance in particular tasks and attended to students' learning strategies.

However, T3 did not consider her 'process assessment' practices in class a success. She considered herself as "focusing too much on finishing her teaching tasks" but "sometimes neglecting to regulate students' learning habits" (T3:8). As T3 argued, the main reason for this was that it took "tremendous energy and patience" for teachers to "constantly check on students' performance and remind them what they should do", particularly when she needed to care for over 100 students. "Too many things need to be done repetitively – but it's my problem, I didn't do well enough", she said (T3:8). From T3's perspective, 'process assessment' was an "art of balance" (T3:8). If teachers have an overwhelming teaching workload and are responsible for large classes, it would be extremely challenging for them to look after every student in their learning progress and strategies.

T4, on the other hand, was unfamiliar with the term 'formative assessment', and I needed to explain the term to her before asking any further questions. Interestingly, her immediate reaction after my explanation was, "Ah! Process assessment!" (T4:6) T4 reported that she had heard of the term 'process assessment' "from other senior teachers" before but found the term "abstract and challenging to implement" in her teaching (T4:6). "How do you assess students during the process", she said (T4:6). Reflecting on her practices, T4 considered her regular bite-size dictation as a kind of 'process assessment' practice. Along with T2, T4

seemed to have perceived formative assessment as regular tests. Such a perception resembled other reports on Chinese teachers' understandings of formative assessment (Chen, 2017; Chen et al., 2013), in which the term 'process assessment' was adopted to describe the continuous tests and grading of students' performances, which largely served a summative purpose rather than a formative one.

In terms of the influence of formative assessment on teaching and learning, T4 believed that 'process assessment' could "supervise and urge students to memorise the important knowledge points of the units" and "help teachers check which knowledge point was not well mastered by students" (T4:7). In other words, she considered that the bite-size tests she assigned could encourage students to remember the language knowledge learnt and assist her in examining students' progress. As T4 argued, bite-size tests could be effective ways to improve students' learning, especially for those students who fell behind, as the tests could "break down the more challenging memorising tasks into manageable ones", which might "boost their confidence" (T4:7). The test results could also inform her about "what aspects of language knowledge students have not memorised" so that she could "practise the knowledge over and over again" (T4:7). From her statements, it seemed that T4 viewed language could be learnt through memorising small components of language and repeated practices. This somehow contradicted what she articulated in her intended teaching objectives, in which applying language in communication was the ultimate purpose of language learning. In the after-observation interview, T4 defended her teaching design with students' level of English:

> Many of my students can't write or speak a complete and correct sentence. I think … they need to accumulate more vocabulary and grammatical rules … so that they can use what they learnt when they encounter a related task. (T4:24)

T4's explanation of her practices indicated that she considered that learning linguistic elements such as vocabulary and grammar could subsequently bring about the development of more sophisticated language skills of students. This could explain why she thought bite-size tests on separate language knowledge could benefit students' learning. Along with her summative-oriented interpretation of formative assessment, T4 placed more attention on the results of the regular tests and pinned her hope on that the pressure of tests and teaching through repetition could improve students' learning. However, without proper task designs

and further instructions on learning from mistakes, students might learn little about how to use language in a context, which she claimed to be her teaching objective.

6.4.3 Experiences related to assessment

The teachers discussed their experiences related to assessment in the interviews. They began with their past professional learning experience regarding classroom assessment. Like the two teachers in School A, both T3 and T4 reported that they had limited academic and professional opportunities in assessment before they became teachers. T3 argued that during her past learning experiences, assessment was largely regarded as "teachers' praise, criticism, and tests" (T3:9), while T4, as a recent graduate in translation studies, reported that she was "not familiar with the different teaching and assessment concepts" until she became a middle school teacher (T4:7). Nevertheless, T3 mentioned that she participated in several teacher education projects held by the local education bureau, which aimed at "introducing teachers to novel teaching and assessment approaches, including pair work, group work, and self-assessment" (T3:9). The opportunity was "opened to selected teachers from different schools in the district", and the teachers, who were possibly recognised as highly skilled teachers in their schools, "were required to pass the knowledge they learnt from the project to their teachers colleagues" (T3:informal interview). She considered these projects "informative and useful" and suggested that she has "incorporated the recommended approaches in her lessons" (T3:10). The observation indicated that T3 had integrated the above approaches in her teaching, which distinguished her lessons from the other three teachers. This might suggest that attending teaching education projects related to assessment could positively impact teachers' everyday practices, particularly for teachers who are agentic in making changes to classroom assessment activities.

The teachers also discussed their present experiences with classroom assessment, particularly the support they received throughout teaching. Both teachers argued that they received few suggestions or guidance on classroom assessment from their teaching resources, such as textbooks, teacher guidebooks, and workbooks. As the two teachers suggested, "the exercises" included in the official teaching resources and "the self-assessment form" at the end of each unit in the textbook were the only "assessment materials" provided (T3:9; T4:8). While the teaching materials offered limited support regarding classroom assessment design, the teachers reported that they often drew on other teachers' experiences regarding classroom

assessment and adopted the engaging or effective assessment approaches they observed in other teachers' classes. Such observations took place within School B and often involved "several teachers attending one teacher's lesson" and "making comments" on his/her teaching after class (T3:6). For example, T3 mentioned an "interesting quick reaction game" she observed in a public lesson, in which the teacher "encouraged students to review the words they have learnt through competitions" (T3:25). T4, on the other hand, stated that the bite-size dictation she often implemented in her class was borrowed from another senior teacher. The teachers' statements revealed that a teacher learning community was also established in School B to promote teachers' professional development. It was unclear in the data whether the community might benefit from novel assessment concepts teacher members brought in, but from the narratives of the teachers, they seemed to regard the experiences as positive ones. Such a community might strongly influence novice teachers, who might depend significantly on imitating experienced teachers' practices.

Apart from their past and present experiences, the teachers were also interviewed about the role assessment played in their future actions. T3 viewed assessment, particularly external tests, as a major influence on the tasks many teachers prescribed in class. As she argued:

> The tasks we prescribed to students were mostly based on the tests they took … Nowadays, the tests are becoming increasingly complex, so we have to focus more on the exercises from the official textbook and workbook because they followed the Zhongkao design. (T3:9)

T3's argument suggested that the design of external high-stake tests, such as Zhongkao, which students would take at the end of their third year, had considerably impacted how teachers chose the focus of their lessons and the tasks assigned in class. Despite the enormous effect of the tests, T3 suggested that she tended to follow the idea that "if students have a good command of language, the test cannot overwhelm them" and tried to "foster students' language skills instead of simply their test-taking skills" (T3:12). She seemed to have prioritised language skills development rather than test-taking in her teaching. T4, on the other hand, was more concerned about how students performed in external tests. Similar to T2, T4 mentioned how "getting a good score in Zhongkao" (T4:25) matters in students' promotion to public high schools and considered it necessary to practise "the skills and knowledge required in Zhongkao-taking through exercises" (T4:15-16). Her reactions indicated that Zhongkao had played a significant role in shaping her teaching and assessment planning.

6.5 Understanding of assessment policies

After discussing the teachers' understandings of assessment, this section explores how the teachers interpreted and interacted with the wider policy context and discovers teachers' enactment of assessment policies at different levels and aspects, including the school policy (section 6.5.1), the Zhongkao policy (section 6.5.2), and the curriculum policy (section 6.5.3).

6.5.1 School B's policy regarding assessment

The two teachers in School B were invited to share their experiences regarding assessment policies from the school level and reflect on how the policies have influenced their work. According to T3 and T4, School B did not establish a specific policy regarding evaluating teachers based on scores, which was distinct from School A. As T3 commented, teachers in School B were "not under immense pressure". However, scores remained an essential factor in School B in terms of accountability. As T4 reported, the school administration valued students' scores, particularly formal examinations, such as mid-term and end-of-term tests. After each formal examination, the administration would analyse each class's test performance and ask the teachers whose classes failed the test to reflect on the reasons for the failure.

School B appeared to have established hidden policies regarding teacher evaluation. Although there was no specific policy that rated teachers as different ranks, the students' test performances were closely supervised by the school administration and were used as evidence of the quality of teachers' work. It was interesting that T3 and T4 reacted to the accountability pressure differently. T3 seemed largely unaffected by the school's evaluation. From her narratives, she retained her skills-focused teaching practices and beliefs and did not recognise a massive burden from the school. Her role as a senior teacher and a management team member might have given her more confidence in following her beliefs in practices. However, T4 seemed to be more troubled by the school's evaluation. She was more concerned with students' test performance and considered it needed to teach to the tests. It was possible that her position on the lower end of the hierarchy in the workplace did not provide her enough confidence to act against the school's expectations or that she needed to establish her reputation through students' test performances in the workplace.

6.5.2 The influence of the Zhongkao reform

Both teachers agreed on the major changes in the English Zhongkao: the increase in the score of the listening-and-speaking test (from 15% to 25%), the removal of vocabulary tasks, and the increase in the score of reading comprehension tasks. The increase in the listening-and-speaking test score has influenced teachers' assessment and teaching of speaking and listening. According to T3 and T4, English teachers and students in School B used an online practice and preparation platform, which was similar to the one School A adopted. The platform provided exercises resembling the English Zhongkao, which involved mock listening-and-speaking tests and could automatically rate students' performance. As T4 commented, the reason for School B to use the online platform was "associated with the increase in the score of listening-and-speaking test" (T4:10). Both teachers reported having assigned mock tests regularly on the platform as homework to students. However, similar to T1 and T2, they both admitted that they did not listen to students' recordings in the mock tests. T3 agreed that "checking on students' performance and providing feedback would be a better approach" (T3:13) yet pointed to her excessive workload as a barrier. T4 considered the automated score as the feedback to students and, therefore, did not check the recordings or explain the tests to students. Students were left alone in reviewing their performances, and it was unknown whether they benefited from the automated test reports.

The two teachers did not mention any changes in their teaching related to removing vocabulary tasks and increasing reading comprehension tasks in Zhongkao. However, their emphasis on the reading sections of the textbook was observable in their classroom practices. Within the nine lessons observed in T3's class, four lessons were distributed for reading-related sections in textbooks and reading exercises, while for T4's class, four out of ten focused on reading or reading exercises. One reason for this phenomenon could be the relatively large proportion of reading-related sections assigned in the textbook (see Chapter 4 materials analysis for more information) and the teachers' compliance with the textbook design. Another reason could be that the teachers found reading a crucial section in language learning and test preparation and considered this section required more attention in class.

The influence of Zhongkao on teaching writing was also discussed in the interview with T3 and further observed in T4's class. According to T3, while she believed that "overusing Chinese in English teaching could impede language acquisition" (T3:6), many other teachers,

including early-career teachers who observed her class, insisted on using a 'sentences translation' approach (as specified in T1's report in section 5.5.2), emphasising the "one-to-one correspondence between English words/phrases and their Chinese translations" (T3:7). From T3's perspective, such a practice "resulted from the writing task design of external exams, including the district's unified examinations and Zhongkao", which "involved too many Chinese instructions" (T3:7). She argued that the design "encouraged teachers to teach translation" and could "inhibit students' creativity in making up sentences in a real-life context" (T3:7). From the observation, T3 did not pre-supply any Chinese sentences for students to translate in their writing. Instead, she set a topic for students to search for information online and required students to write according to their research. Her approach differed sharply from the 'sentences translation' approach and aimed to foster students' writing skills instead of merely grammar and lexical skills.

On the other hand, the classroom observation showed that T4 relied on Chinese translation in her teaching and followed the 'sentences translation' approach in teaching writing. In her writing class, after reviewing the words, phrases, and passage structure of a sample article provided by the writing section in the textbook, T4 assigned students a writing task which echoed the topic of the unit, along with five Chinese sentences, which were the translation of a sample article. She asked students to translate the sentences into English and then integrate the five sentences into an article with proper conjunctions. According to T4, this approach was a "routine" for teaching writing and could "scaffold students to write up a complete essay" (T4:23-24). However, the approach could only improve students' grammar and lexical competence and might have merely created an illusion that students could write well: "If they are always translating the language, they are not using the L2 for communication" (Harmer, 2007a, p. 49). If translation instead of writing was practised, students could not develop strategies that helped them compose an article independently. It is questionable whether such an approach could lead to better writing skills of students.

6.5.3 The influence of the curriculum standards

The influence of the latest English curriculum standards on the teachers' teaching and assessment was also investigated. T3 reported that she "did not look into" the 2011 curriculum standards and "did not use them as a reference" for her teaching and assessment (T3:14). According to T3, at the beginning of her teaching career in the early 2000s, the

school provided her with a copy of the 2001 curriculum standards, the trial version of the latest curriculum. Although T3 considered "the language level descriptors" enlisted in the curriculum standards "helpful" at the beginning of her career, as they helped "portrait the levels" students should reach, she admitted that "the Zhongkao remained a powerful impact" on her teaching (T3:14). The assessment guidelines in the curriculum standards, as T3 argued, were "inconvenient to operate" since they were "general" and "difficult to meet every teacher's need" (T3:14). T4 reported having read the 2011 curriculum standards yet considered the document "has little impact" on her teaching and assessment (T4:11). From her perspective, the general descriptions of the language levels and the assessment principles were "not practical enough", as they "could not be applied directly" in teaching her students (T4:11). She considered "the lesson design and advice from experienced teachers" the primary reference for her teaching and assessment planning (T4:11). It seemed that the curriculum standards have a rather feeble effect on both teachers' teaching and assessment.

6.6 Summary

This chapter summarises the two teachers' classroom assessment activities, their understanding of assessment, and their understanding and enactment of the assessment policies influencing their context. According to the data, the two teachers shared some features in their teaching and assessment beliefs regarding how language skills should be emphasised but diverged concerning actual practices. T3 adopted a contextual approach to language teaching and assessment, in which language is taught and assessed in a context so that students can connect with what they learnt. She implemented both open and closed questions to elicit students' responses, focused on teaching both the form and meaning of English language, and offered feedback from both task and learning strategy levels. She embraced a more student-centred attitude in design assessment tasks, as she attempted to employ encouragement and prompts to promote students' learning and took an open attitude towards using student-centred assessment activities. T4, on the other hand, adopted a more behaviourist-oriented approach to teaching and assessment. She relied more on closed questions and tasks in classroom interactions, tended to concentrate on teaching linguistic knowledge, and provided restricted feedback on the task level. She incorporated a more teacher-centred attitude in her assessment design, as she was more concerned with how she wanted students to perform instead of how their learning could be improved through classroom activities and feedback.

The data also showed that the two teachers held different levels of understanding of formative assessment. T3 interpreted formative assessment as helping students foster better learning strategies so that they could regulate and monitor their studies more efficiently. Even though she reported not having implemented formative assessment successfully in her classes due to the large class size and the resultant heavy workload, she has touched the core of what formative assessment means, despite calling the concept a different name. In contrast, T4 regarded regular bite-size tests as representations of formative assessment, which resembled the conceptions of T1 and T2. She adopted a relatively narrowed conception of formative assessment, in which continuous summative testing was considered formative in nature.

Inquiries into the teachers' experiences related to assessment and the policies they worked with offered us another perspective to understand their assessment activities and understandings. Both teachers stated that they hardly received professional education regarding classroom assessment before they became teachers and that the official teaching materials, including textbooks and guidebooks, did not offer sufficient assistance in planning classroom assessment activities. Tests continued to be a significant factor used to evaluate teachers and have a major influence on the teachers' teaching and assessment design, particularly novice teachers. Zhongkao defined the future goals of the teachers, and the assessment guidelines in the curriculum were largely ignored. Despite the test-oriented context, there were encouraging phenomena. Being involved in teacher education projects introducing novel assessment approaches has brought changes to T3's classroom practices. Teachers in School B were also participating in the local teacher learning community, in which assessment approaches that could effectively engage students could be shared and developed.

Chapter 7 Discussion

The previous chapters analysed the textbook, teacher guidebook, the Zhongkao test design (Chapter 4), the four teachers' assessment activities, understanding of assessment, and understanding of assessment policies (Chapter 5 and 6). The analysis aims to answer the three research questions of this study, which look at the teachers' assessment activities, their assessment beliefs, and their enactment of the local assessment policies. Adopting bounded relativist and constructionist positions, the analysis adopted a hybrid inductive and deductive approach of interpretation by both working bottom-up to recognise the details in the dataset pertinent to the teachers' assessment practices, their understanding of assessment and assessment policies and contextual factors influencing teachers' practices and thinking, and exploring these social phenomena observed through the lens of the theories reviewed, including language theories, assessment theories, and theories regarding teacher agency and policy enactment. Through these theoretical lenses, the assessment practices, assessment understanding, and enactment of assessment policies of the teachers can be better comprehended.

The discussion centres on three aspects, which intends to answer the three sets of research questions. Section 7.1 addresses the first and second sets of research questions, reporting the assessment activities in the two schools and the language learning and assessment beliefs behind teachers' choices. Section 7.2 addresses the first and second set of research questions in a more in-depth way, discussing the purposes underlying the teachers' assessment activities. Section 7.3 addresses the third set of research questions, analysing the teachers' policy enactment by discussing the assessment context and how it influences teachers' work. Following the discussion, I propose a framework (section 7.4) for understanding classroom assessment activities and teachers' assessment beliefs in this Chinese context. Section 7.5 summarises the discussion chapter.

7.1 Assessment activities in the two schools

Chapter 5 and 6 document three types of assessment activities in the two schools: oral tasks, written tasks, and student-assessed activities. This finding matched Rea-Dickins's (2001) report on the assessment activities observed in English classrooms in the UK, which involved written tasks, informal classroom conversations, and student-initiated assessment

tasks. The teacher participants in the two schools shared some commonalities but also showed differences in their classroom assessment design. Oral tasks were adopted most commonly in both schools. All teachers adopted display questions, which is consistent with Walsh's (2011) description of non-communicative classroom discourse, to evaluate students' responses, while one teacher (T3) was observed to use referential questions to extend students' contribution and explore what students could do in using the language knowledge, which was a feature of communicative teacher talk in Walsh's (2011) report. The teachers tended to assess students' reading skills and their memory and understanding of vocabulary and grammar through oral tasks, which corresponded with the focus of the textbook and the Zhongkao, as reported in Chapter 4. Written tasks were adopted less frequently in both schools. All teachers implemented written exercises from textbook and workbook, which had been fashioned after the test items in the mid-term exams, the final exams, and the Zhongkao, to assess students' language knowledge and foster their test-taking skills. Student-assessed activities were the least common strategy in both schools. Teachers in School A (T1 and T2) adopted a relatively negative attitude towards such activities, as they felt either distrustful of the students' capacity to assess or not confident in offering students more freedom in taking the assessment initiative. Teachers in School B (T3 and T4), on the other hand, appeared to be more confident and active in encouraging students to take part in self-marking and peer discussion, which might allow more opportunities for students to reflect on their performances (Leung, 2020).

The study showed that the teachers embraced different language learning and assessment views. T1 and T2 considered teaching and assessing specific language elements as their objectives and relied on closed questioning and tasks to examine whether students could reproduce the knowledge they transmitted. They seemed to treat language as a behaviour that could be formed through mimicry and memorisation (which allowed imitation to happen), repetitive drills (which reinforced a behaviour), and encouragement (which served as rewards), as advocated by proponents of behaviourism such as Brooks (1966). Such views sat within the behaviourist paradigm, in which learning is believed to be based on stimuli and desired responses (Brown & Abeywickrama, 2019; Lightbown & Spada, 2013). It is not surprising that the two teachers expected students to provide 'accurate' answers to their questions and relied heavily on evaluative feedback to strengthen students 'desired behaviours'. The data showed that T2 recognised the benefits of active student participation, as she commented, "a sense of getting involved could … positively influence students'

learning motivation" (T2:19). However, she seemed to define 'active participation' as students answering her questions actively instead of students taking an active part in the construction of learning activities. Students remained passive receivers in the activities. These findings are, to a certain degree, similar to other studies which focused on Chinese secondary school English classrooms, in which linguistic knowledge was given priority in class, and a limited range of assessment types and feedback techniques was available (e.g., Gu, 2014; Zheng & Borg, 2014).

Compared with T1 and T2, T3 adopted a more usage-based view of language learning and assessment, as demonstrated in Chapter 6. She considered her teaching objectives as instructing students to learn how to use the language knowledge and implemented both closed and open questions and tasks. From the observation, T3 assessed both 'if' and 'what' the students could do, which matched Torrance and Pryor's (2001) description of convergent and divergent assessment approaches. Language in her class was viewed as the knowledge that needed to be "learnt and taught in a context" (T3:6) and required active individual sense-making. She also regarded self-regulation and self-reflection as critical and sought to create more spaces for students to take more responsibility in classroom assessment. The analysis showed that T3's approach to teaching and assessment could be considered constructivist, as the active process of thinking, rather than merely absorbing knowledge, was encouraged (James, 2008). Similar to the two teachers in School A (T1 and T2), T3 assessed students' memory and understanding through closed questions and described Zhongkao-simulating tasks as necessities in teaching. This suggests that assessment activities that bear different theoretical implications can be housed under the same roof when different purposes need to be accomplished, for instance, consolidating communicative skills, strengthening the memory of a particular language knowledge, and fostering test-taking skills. This also suggests that strict one-to-one relationships might not exist between teachers' assessment activities and beliefs when external factors shape teachers' decision-making, which is consistent with Borg's (2018) and Li's (2013) argument.

T4 adopted seemingly contrasting language teaching and assessment beliefs, which mixed a constructivist aim with behaviourist strategies. In her espoused beliefs, T4 expressed aspirations for leading students toward communicative-oriented language learning and instructing them to use the language knowledge they learned. Such beliefs might result from the changing curriculum ideas or the influence of T3, her mentor and department leader.

However, her actual practices bore a strong resemblance to the grammar-translation method, in which students were presented with word lists and grammar rules and were required to make use of the words and rules in translation exercises (Brown & Lee, 2015) instead of the contextual approach proposed by T3. From the interviews, it was inferred that T4 believed that understanding the declarative knowledge of English, such as grammatical rules and vocabulary, could subsequently lead to the procedural knowledge of using the language (Paradis, 2009). It cannot be denied that the grammar-translation method is attractive to many teachers, for it does not require teachers to have specialised teaching skills, and tasks related to grammar rules and translation can be easily scored (Brown & Lee, 2015). However, overfocusing on vocabulary and grammar and overusing translation exercises means that students focus too much on the linguistic elements and English equivalents of Chinese (and vice versa) and have few opportunities to practise communicating effectively with English (Harmer, 2007a). This also suggests that inadequate knowledge and methods of language teaching and assessment might result in limited abilities of teachers to enact their beliefs (Borg, 2018; Levi & Inbar-Lourie, 2020).

7.2 Teachers' purposes of assessment

Besides the theoretical implications embedded within the teachers' assessment activities, the purposes behind the teachers' assessment design should be further explored. The findings revealed multiple purposes in the teachers' classroom assessment activities. To begin with, many of the observed assessment activities were conducted by the teachers to guide their teaching instructions at later stages (I refer to this purpose as 'for-instruction purpose' in the following discussion), which is similar to the findings of Cheng et al. (2004). The teachers from both schools mentioned that they elicited students' oral and written responses through classroom assessment to monitor students' progress and plan their instructions accordingly. From the analysis, the teachers tended to view assessment activities, particularly oral and written tasks, as making judgments on students' performance and based their feedback on whether the students' responses matched their expectations. They assessed so that they could make instructional decisions according to the information obtained about students' attainment and address students' problems with further elaborations on the knowledge that they considered students failed to absorb. While guiding teachers' instructions is an important function of classroom assessment, the role of students and how students can be led to their learning goals, which should be the focus of formative assessment (Assessment

Reform Group, 2002), did not seem to receive much attention when the teachers designed these assessment activities. The lack of feedback on how students could improve, as reported in both cases, also demonstrated that helping students' learning was not placed at the centre of teachers' classroom assessment in many instances.

Other than for-instruction purposes, some of the teachers' assessment activities served more for-learning purposes. The most typical examples are identified in T3's classroom, where she incorporated feedback about learning strategies along with her task-based feedback. Students in her class were instructed to develop what she called 'good learning habits', such as previewing, reviewing, and taking notes, so that they could self-manage and self-regulate their own learning in and after class. Learning in her class was viewed as more than accomplishing a list of tasks – the strategies to attain the learning goals were also focused on (Hattie & Timperley, 2007). The teachers referred to other assessment techniques which they believed to have made students more agentic in learning. One technique that the teachers frequently mentioned was praise and encouragement, which was also reported by many studies conducted in the basic education sector in China (for example, Sun, 2016). Arguably, praise and encouragement can have affective benefits that potentially help students gain more confidence and motivate them to participate more actively in classroom activities (Meng, 2009). However, the analysis showed that the praise and encouragement teachers provided typically included little detailed guidance on further improvements, such as task-related or strategy-related information, and might have little impact on students' understanding of the tasks (Hattie & Timperley, 2007; Zhang, 2013).

Three teacher participants (T2, T3, and T4) also reported other assessment approaches, such as self-marking and peer discussion, as having a positive influence on students learning. They considered that these techniques had created a more vibrant learning atmosphere by encouraging students to reflect on their performances and learn from others' strengths and weaknesses. These student-assessed activities are reported to have the potential to empower students to deepen their understanding of knowledge, as students were given the opportunities to develop their capacities for self-assessment and their understanding of the learning goals and criteria (Gardner, 2012). However, it should be noted that while students in the two cases were instructed to play seemingly active roles in these classroom activities, their movements were essentially teacher-controlled. In most cases, students were left alone to conduct self-marking or discussion tasks without further guidance from the teachers. What

can be described as largely student-controlled assessment activities are the peer assessment moments identified in the two novice teachers' (T2 and T4) classes, in which students detected and corrected their peers' mistakes without the teachers' instructions. Yet, there was little evidence suggesting that the two teachers made use of students' enthusiasm and made further changes in their assessment design. As Liu and Xu (2017, p. 27) argued, students' involvement in formative assessment "tend to be low without the teacher's careful planning of the activity, thoughtful training of students as peer assessors, and sustained support throughout the process". If the students are to be truly active agents in the assessment process, the teachers need to supply more ongoing advice on how students should provide feedback to others and how they can learn from their learning experiences.

Apart from the purposes mentioned above, the analysis also revealed that teachers conducted classroom assessment for maintaining classroom discipline, which was a purpose not directly related to teaching and learning. Three teachers (T1, T2, and T4) expected students to stay focused in class and respond to teachers' questions when asked and reported that oral assessment activities could assist them in achieving these goals. One teacher (T2) even expected students to perform at the same time in classroom assessment tasks so that she could better manage the classroom and therefore considered written tasks a better approach to evaluate students. The teachers' emphasis on classroom discipline corresponded with how students were arranged to sit in the classrooms, in which whether the arrangement could contribute to effective discipline was especially taken care of. This corresponded with Brown and Gao's (2015) report on Chinese teachers' conceptions of assessment, in which assessment was considered a useful approach to maintain order and discipline in classrooms. The reason why these assessment tasks could serve such a purpose might lie in the classroom culture of the context, where teachers are generally regarded as authoritative roles in classrooms, and students should behave as teachers instructed (Cheng & Ding, 2021). As argued in Chapter 5 and 6, the teachers' rationales for maintaining classroom discipline should not be denied, as in a large class, maintaining classroom discipline could bring benefits to the construction of a more organised and efficient condition in which learning could better occur (Lopes & Oliveira, 2017). However, it should also be noted that raising questions to students constantly might only correct students' behaviour temporarily but has little long-term effect on improving their learning. Other approaches should be adopted to help students make substantial changes in their learning behaviours and strategies.

Analysis of teachers' purposes of assessment showed that there are gaps between the teachers' assessment activities and the formative assessment principles. The teachers mainly designed assessment from a teacher-centred perspective, for instance, how they should plan their teaching and manage their classrooms, instead of a student-centred one. All but one teacher (T1, T2, and T4) paid less attention to how students' learning strategies could be developed from a long-term perspective and seemed to consider that ongoing judgements and elaborations on knowledge points could effectively improve learning. Teaching and assessment were also viewed as two separate procedures instead of an entity of effective planning: assessment followed the instructions on the knowledge points, and evaluative feedback with further instructions on the knowledge points followed the assessment outcomes. Such a phenomenon might explain how these teachers interpreted formative assessment as ongoing tests, praise and encouragement. They understood formative assessment as producing positive and negative results, which triggered the next step of their instructions. Such approaches and beliefs could be classified as restricted forms and interpretations of formative assessment (Carless, 2011). One teacher (T3) took a further step in recognising the importance of directing students to foster improved learning strategies and encouraging students to take more autonomous roles in learning and assessment. Such an approach has more substantial potential for students' learning, as students were given the opportunity to learn how to learn and develop self-regulation strategies, which are essential skills for both learning at present and life-long learning (Black & Wiliam, 2009).

7.3 Assessment context

The study shows that the teachers' assessment activities and understandings were deeply influenced by the assessment context they worked in. This section discusses how the assessment context impacted teachers' work using Priestley et al.'s (2015) teacher agency model. The model explains teachers' social actions from the iterational, projective, and practical-evaluative dimensions. The iterational dimension (section 7.3.1) looks at teachers' prior educational and professional histories. The projective dimension (section 7.3.2) refers to teachers' future aspirations about their work. The practical-evaluative dimension (section 7.3.3) looks at teachers' day-to-day working environment. Teachers' past experience, future orientation, and present engagement are jointly considered in terms of how they inform teachers' actions. Braun et al.'s (2011) four contextual dimensions (situated, professional, material, and external contexts) are also used to support the discussion.

7.3.1 Iterational dimension

For the purpose of this study, the teachers' prior educational and professional experiences regarding assessment are given special attention. All teachers reported that they had not received professional education opportunities regarding classroom assessment before they became teachers. Therefore, it is not surprising that the teachers had a relatively narrow range of assessment toolboxes, in which closed oral and written tasks comprised a significant share of the observed assessment activities, while other student-centred assessment activities were less frequently adopted. They were given few instructions on what should be done in classroom assessment and had little exposure to successful examples of assessment activities that promoted students learning from both task and learning strategies perspectives. The finding resonates with Gu's (2014) report that the secondary English teacher in a top school in Beijing, who received no training at all in assessment, tended to resort to fixed-response assessment tasks and offered mainly evaluative feedback, which indicated right or wrong, and descriptive feedback, which included item-by-item explanations of the tasks. Only one teacher (T3) participant reported to have participated in teacher education projects related to novel assessment approaches during her teaching career (details will be discussed in section 7.3.3.2), and she happened to be the one who conducted student-assessed activities the most often and included suggestions for developing students' learning strategies in her feedback. This might suggest that limited professional education opportunities in assessment could restrict teachers' capacities to master more diverse assessment approaches and that the professional context, as mentioned in Braun et al. (2011), has an impact on how teachers enact the assessment policies.

The data also showed that many teachers seemed to hold or were given the idea that teaching is not interconnected with assessment, or teaching matters more than assessment in class. One teacher (T1) reported that the teacher education programmes she attended heavily focused on pedagogical approaches instead of assessment. From several teachers' (T1, T2, and T4) narratives, it was also noticed that they seemed to consider assessment and teaching as separate educational procedures – classroom assessment is conducted at the end of a teaching period, and its results are used to adjust their teaching. This suggests that such ideas might prevail among teachers and local teacher educators in the context under investigation. Such phenomena are common in many education settings, where teacher education gives little prominence to assessment literacy and discusses teaching practices separate from

assessment practices, even though assessment has always been involved in teachers' work (Gotwals & Cisterna, 2022; Leung, 2014). They could subsequently lead teachers to design classroom assessment based on what they thought might better facilitate teaching, for instance, judging students' performances through assessment tasks and using the information to plan their instructions or using assessment tasks to manage classroom discipline for a better learning environment. The for-learning purposes are not entirely neglected in their assessment designs but are not considered as the central purpose of assessment either.

The analysis also showed that prior professional education experience could influence teachers' interpretations of formative assessment. Among the four teachers, three (T1, T3, and T4) referred to formative assessment as 'process assessment' (as explained in section 5.4.2, pp. 112-113), which was a term they learnt from teacher education projects or senior teachers, while one teacher (T2) claimed that she had never heard of the term formative assessment. All but one teacher (T1, T2, and T4) considered formative assessment as assigning assessment tasks continuously to students throughout the teaching process and using the assessment results to guide their teaching and students' learning further. The findings partly resonated with Chen's (2013) study, which discovered that English teachers in Chinese universities adopted the term 'process assessment' to describe formative assessment and foregrounded the timing and frequency of continuous assessment throughout a certain teaching period. What is new in this study is that one teacher (T3) accepted the term 'process assessment' but interpreted the term as helping students develop better learning strategies. Interestingly, the term 'process assessment' seemed to have received wide recognition in China's secondary and tertiary education sectors. Judging from the teachers' narratives, even the teacher educators who designed professional development projects for them accepted this term. It is unknown how the teacher educators explained formative assessment or process assessment to the teachers, but from the analysis, they have not successfully passed down what really matters in formative assessment to many teachers in this context. These professional development projects should perhaps review their courses to ensure that the definition of formative assessment is accurately conveyed.

7.3.2 Projective dimension

The teachers' aspirations about the future also shaped their present practices in classrooms. To be more specific, what the teachers expected to achieve in the short term and long term

could impact how they planned their classroom assessment. The analysis showed that the teachers' assessment activities and what they believed was the best for their students were strongly related to the high-stake summative test, namely the Zhongkao. The observations revealed that the teachers' classroom assessment emphasised assessing vocabulary, grammar, and reading skills, which matched the focus of the Zhongkao design. The interviews also demonstrated that summative tests, such as unit tests, final tests, and Zhongkao, have strengthened their beliefs that learning, teaching, and assessment should focus on the high-stake tests their students would take. Such phenomena are widely reported in secondary education sectors of different contexts (for example, Ali & Hamid, 2020; Ma & Bui, 2021; Mansell et al., 2009) and call into question what purposes of assessment should be prioritised – ensuring students' achievement, evaluating the quality of schools, or selecting students into different institutes.

As reported in section 4.2 (see p. 84), the Zhongkao guidelines describe Zhongkao as both an achievement test that judges whether students could obtain a middle school qualification and a selection test that decides whether a place in high schools could be offered to students. It is designed to serve two distinct purposes. However, Zhongkao seemed to be viewed more as a selection test by the teachers from the analysis. The teachers emphasised the consequence of students not receiving good scores in Zhongkao as not having access to public high schools and deemed it necessary to teach to the test. They were also more concerned about what was tested in Zhongkao and tended to teach the language knowledge and skills that received more attention in the test rather than what was prescribed by the curriculum. The selection purpose appeared to outweigh the qualification purpose, as being selected into public high schools had higher stakes.

It also seemed that the qualification and selection purposes of Zhongkao conflicted to a certain extent with each other. The selection function of the test was based on students' scores, yet good scores did not guarantee students' competence in the communicative skills required in the curriculum. The findings corresponded with Qi's (2005) report on the divergence between the intended use of assessment and its actual use in the Chinese secondary context, in which the functions of selection drove teaching to the test, and the function of promoting changes failed to achieve the desired outcome. When there is a mismatch between the scope of the curriculum and that of assessment, the impact of assessment on promoting better teaching and learning can be restrictive, and the curriculum goals might be interpreted

narrowly (Harlen, 2007). The teachers might also develop a strong interest in the reliability of assessment, which refers to how an assessment can provide consistent evidence of students' progress (Mansell et al., 2009) due to the role high-stake tests played in their work experience. They might tend to resort to objective, reliable, and closed tasks, which took the form of the tasks in the high-stake tests, as their classroom assessment activities, despite that the interpretation and use of the information achieved from these tasks matter more in classroom assessment (Stobart, 2012).

Nevertheless, the teachers' concern about students' scores and promotion to public high schools remained a realistic issue. The promotion rate of students in Shenzhen from middle schools to public high schools is long considered low by the public as a result of limited public high school resources. The annual promotion rate was not made public by the local education bureau, as it was considered inappropriate because such a move might exacerbate stress among educational stakeholders and unfair competition (Shenzhen Admission and Examination Office, 2020). However, in another announcement made by the bureau, it was proposed that by 2025, the promotion rate will be increased to 56% by building more public high schools (Shenzhen Education, 2020), suggesting that the promotion rate is indeed low as the public perceived. There are historical reasons behind this phenomenon. As a young city founded as a Special Economic Zone in the 1980s after the release of the Reform and Opening-up Policy, Shenzhen has been known as a migrant city that offers numerous work opportunities for people around the country. By 2020, Shenzhen was a first-tier city in China, which ranked third for the overall gross domestic product in the country, trailing only Shanghai and Beijing (National Bureau of Statistics, 2020). During the past 40 years, the city has experienced explosive growth in population and economy. However, along with its prosperity, there was an incompatibility between the growing population and the educational resources. The short history of the city and the inadequate investment in education has led to the current situation (Jia, 2013).

From this situation, it can be inferred that the promotion rate for the student group under observation was not optimistic, as they took the Zhongkao in July 2022. It is unclear whether the situation will improve in the next couple of years if the promotion rate is improved as promised. Admittedly, entering a public high school is a decent opportunity for students to pursue further education. For fifteen-year-olds, leaving the school system is usually not a good choice; entering a private high school means an enormous expense for some families;

vocational schools are generally considered places where ‘bad students’ go to, especially for middle-class parents (Lin, 2015). The impact from the situated context, where test results were greatly valued, and the external context, where pressures and expectations from other educational stakeholders came from, shaped the teachers’ enactment, which was consistent with Braun et al.’s (2011) work. Understandably, the teachers took the scores seriously, as good scores could at least guarantee better educational opportunities for their students. When the educational resources in the context remain limited, compromises have to be made among validity, reliability, fairness, teachers’ time, and expertise (Harlen, 2007). For this case, working for the test has become a shared aspiration among different stakeholders. External systemic changes are needed so that teachers can transform themselves as agents of assessment reform.

7.3.3 Practical-evaluative dimension

The teachers’ assessment activities and understandings were also influenced by the factors in their day-to-day working environment. The analysis identified factors from three levels: the classroom level (section 7.3.3.1), the school level (section 7.3.3.2), and the assessment policy level (section 7.3.3.3).

7.3.3.1 The classroom level

Two factors at the classroom level showed up as affecting teachers’ assessment designs and conceptions. The first one is the large size of the observed classes. The analysis indicated that, in all four classes, the class size was a crucial factor in driving teachers to grip discipline as a tool to establish what they considered a good learning environment. It encouraged all four teachers to resort to more manageable assessment activities, such as written tasks, so that they could easily implement the tasks and regulate students’ behaviours. The analysis also suggested that the class size might have limited teachers’ assessment choices, as their relatively narrow scope of assessment knowledge could not assist them in arranging more student-centred assessment activities or offering more constructive feedback to a large group of students.

Large class size has been reported as having a negative impact on cultivating a supportive and cooperative learning environment, as teachers found classroom management in a large class setting physically and mentally challenging (Liu et al., 2021; Yan, 2015). It resulted

from limited available teachers and classrooms, reflecting local constraints from the material context, as Braun et al. (2011) argued. Teachers might consider it difficult to know all students' progress in a large class unless the students submitted for assessment and might turn to frequent examinations as a resolution, in which the focus of assessment might not align with what the curriculum instructed (Biggs, 1996; Sadler, 2010). However, there are potential ways in which these issues might be tackled. For instance, teachers could rethink the aim of their teaching, that is, whether they were simply teaching discrete knowledge points or teaching a higher level of understanding, and reconsider the alignment between their teaching and assessment (Biggs, 2012). They could also develop their knowledge regarding assessment theories, commit more to conducting assessment for students' learning, and use technology to reduce the constraints of large classes (Xu & Harfitt, 2019).

Another factor within the classroom was the students' language level, which could determine the difficulty and complexity of the teachers' assessment activities. The teachers from the two schools provided different descriptions of their students' language levels: teachers in School A (T1 and T2) reported that only a few students could respond well to their questions and tasks, while the majority of the class were having trouble learning English; teachers in School B (T3 and T4) stated that they have relatively more students who were high-performing and fewer students who performed poorly. Such background information might be reflected in the teachers' teaching and assessment, as teachers needed to select classroom activities of an appropriate level of difficulty according to their students' level. From the analysis, teachers in School B (T3 and T4) appeared to adopt more student-centred assessment activities and hold more open attitudes towards these activities than teachers in School A (T1 and T2). It might be the case that students in the observed classes in School B were more capable of completing tasks that required more advanced language skills and self-regulation skills, such as self-marking, discussion, and writing booklets, while students in the observed classes in School A did not have sufficient skills, or their teachers considered them not having the skills, needed in these tasks.

While much of the literature discussed teachers' pedagogical and assessment knowledge as the key factor in constructing a communicative and collaborative classroom (for example, Clark-Gareca & Gui, 2019; Zheng & Borg, 2014), it is often ignored that teachers need to base their teaching and assessment design on students' level, and some activities recommended by the constructivist and socio-cultural language teaching theories might be

challenging for students with lower language proficiency. However, this does not mean that constructivist and socio-cultural theories cannot be enacted with low-proficiency students. Teachers could still concentrate on students' mental processes in learning and guide them to demonstrate their thinking in actions (James, 2008). Students' first language could be utilised at an appropriate level to assist their task completion, and the language knowledge needed for the tasks should be effectively delivered to ensure that students are given enough support to complete the tasks.

7.3.3.2 The school level

The school level witnessed several factors that impacted the teachers' work. One factor was the accountability mechanism the teachers experienced from the school administration. The analysis showed that both schools emphasised students' test performance to monitor whether students have made progress in their learning. The only difference was that School A established clear teacher evaluation policies, which ranked teachers based on students' test results, while School B had a hidden policy regarding teacher evaluation, in which teachers were not ranked, but their work was judged by the test outcomes of their classes. Three teachers (T1, T2, and T4) regarded the accountability mechanism as either pressure or incentive, which encouraged them to focus more on test preparation during teaching. This corresponded with research done in many contexts, where accountability played a substantial role in impacting how teachers linked teaching to testing (for example, DeLuca et al., 2021; Hui et al., 2017). Arguably, accountability is not entirely negative. Hargreaves (2009, p. xv) mentioned that "the purpose of accountability … is also to improve performance by examining its impact, measuring quality and results and spurring people on to achieve even higher standards and greater improvement in the future". However, it cannot be denied that the malpractice of accountability needs to be carefully avoided, as teachers' agency in pursuing changes and embracing new teaching and assessment approaches might be eroded by the heavy accountability mechanism (Priestley et al., 2015)

Another factor was the assessment training available to the teachers. Three teachers (T1, T2, and T4) reported that they received no professional learning opportunity regarding assessment during their teaching career in the schools, while one teacher (T3) stated that she was offered the opportunity to attend assessment-related projects as a selected teacher. It seemed that the context had employed a cascade model for teachers' continuing professional

development, which "involves individual teachers attending 'training events' and then cascading or disseminating the information to colleagues" (Kennedy, 2005, p. 341). Such a model is widely implemented in many contexts for its cost-effectiveness but is also criticised for its limited effect in bringing about large-scale changes at the classroom level, as the knowledge and techniques promoted in the teacher education programmes might be diluted when it is passed down to local teachers (Hayes, 2000; Turner et al., 2017). From the analysis, it could be seen that T3 benefited from the assessment-related teacher education project, as she had successfully adopted various student-centred assessment techniques and understood formative assessment from a constructivist perspective. However, it was unclear whether T3 had managed to pass the knowledge she obtained from the project to her colleagues, as the observation of teaching and assessment of T4, who was T3's mentee, did not demonstrate ample evidence that T4 has followed T3's practices and understandings of formative assessment precisely. The context has not offered a wide range of teacher education programmes for promoting assessment knowledge either and relied on the cascade model to disseminate the knowledge. Such an effort was far from enough to effect more substantial shifts in teachers' activities and understandings of assessment.

A third factor was the teachers' professional learning community identified in the two schools. Teachers from both schools reported having participated in regular teacher meetings or open lessons with their colleagues to observe, discuss, and share each other's experiences. The novice teachers in the two schools also stated that the schools' mentoring systems had benefited their professional development. Such established practices in the two schools demonstrated that despite limited external support, the teachers were actively involved in continuous learning, which allowed them to collaborate, reflect, experiment, and engage (Näykki et al., 2021). However, a potential problem with these practices might lie in the fact that the teachers lacked the channels to learn about advanced theories in language teaching and assessment and to seek advice from external experts, who could provide them with specific guidance on how they could design classroom assessment in a more for-learning way. Notably, for novice teachers in this context, the most reliable reference they could depend on was the senior teachers' guidance, regardless of whether such guidance might actually enhance students' learning or build their abilities for sustainable improvement. Without new knowledge of teaching being introduced to the community, the traditional concepts and methods might be retained and passed down without being questioned. If the teachers' assessment designs and conceptions are to be advanced more effectively and

efficiently, external professional assistance should be introduced and cooperate with the pre-existing established teacher learning community to pour in new ideas and to facilitate supervised teaching and reflections. But again, if the external experts are not aware of the local constraints, it is possible that they could not provide practical advice that the teachers could easily adopt. Communications between the experts and the teachers become significant in negotiating appropriate assessment approaches that match the local needs.

7.3.3.3 The assessment policy level

The analysis illustrated how assessment policies are reflected in the Zhongkao, the textbook, and the teacher guidebook teachers work with and explored the teachers' assessment activities and understandings under the guidance and impact of these materials. The Zhongkao policy is discussed first due to its heavy influence on the teachers' work. The Zhongkao was designed to test the four language skills, vocabulary, and grammar in a listening-and-speaking test and a written test, of which the impact on the teachers' teaching and assessment activities was demonstrated in the analysis. The listening-and-speaking test has driven teachers to assign more listening and speaking assessment tasks for students to complete in and after class, as reported in Chapter 5 and 6. However, the teachers offered only limited feedback on students' listening and speaking performance due to the enormous workload and inadequate time and energy. Automated scoring software was relied on as the feedback provider outside of the classroom, and scores were the only available feedback when students used the software. While an automated scoring system could benefit students' learning by providing efficient feedback, increasing motivation and independence, it might also mediate teachers' instructions in a counterproductive way, leading students to focus on scores only and creating more work for teachers (Wilson et al., 2021). Moreover, as shown in the analysis, the teachers might also consider scores as sufficient feedback for students, who might miss the opportunities for detailed feedback. The teachers need to realise that automated scores are insufficient to promote further improvement in students' learning. Detailed feedback for individual students is needed to boost their language proficiency.

Another impact of the listening-and-speaking test was that the teachers emphasised the accuracy of students' responses, while fluency was somewhat neglected. Such a phenomenon might result from the high stakes of Zhongkao. For a highly competitive selection test such as Zhongkao, a few more marks might make a massive difference to the

kind of schools students get admitted to. In other words, every score matters to students. As a result, the teachers might spend tremendous efforts teaching students how to avoid making mistakes in tests, hoping that students get fewer score deductions. The restricted feedback provided for students' oral performances also mainly concentrated on grammatical and lexical accuracy. The "flow, continuity, automaticity, or smoothness of speech" were given less attention (Koponen & Riggenbach, 2000, p. 6). Arguably, the role of grammar and vocabulary is central to language use and communication and deserves much attention (Tavakoli & Hunter, 2018). However, over-emphasising accuracy might result in a lack of fluency, as students might hesitate and pause too much to prevent making mistakes (Skehan & Foster, 1999). More often, students might adopt a limited participation strategy – "the less they say, the fewer mistakes they make" (Kerr, 2017, p. 8) – which matched what was observed in the four classrooms and reported by the teachers. Highlighting accurate language uses rather than indicating errors and focusing more on the content of students' utterances might help to reduce students' anxiety and increase their participation (Kerr, 2017).

The written test, which comprised reading comprehension tasks, two cloze tasks, and a writing task, also significantly affected the teachers' teaching and assessment. Both reading comprehension and cloze tasks consisted of multiple-choice and gap-filling tasks, which were closed-formatted, easy-to-score, and considered highly objective and reliable (Nusche, 2016). These tasks, assessing students' reading skills and mastery of vocabulary and grammar, accounted for 60% of the total score and received close attention in the teachers' teaching and assessment design. The teachers were observed to spend a much more considerable amount of time teaching and practising reading skills, vocabulary, and grammar, and the written assessment tasks assigned to students primarily focused on these areas and followed the Zhongkao test design. Arguably, these tasks attempted to assess different language skills and knowledge of the students integratively and might have beneficial effects in promoting students' mastery of different skills and knowledge (Oller, 1979). However, the lion's share of the test design given to assessing reading, vocabulary, and grammar had driven teachers to spend more time teaching and assessing these areas, while other aspects of language were given insufficient care, which corresponded with Gu's (2014) report. It was possible that the Zhongkao designers had higher expectations of students' receptive skills and language knowledge instead of productive skills or that it was simply more convenient to test these aspects of language.

The writing task accounted for 15% of the total score and required students to write an 80-word article following the given instructions. From the analysis, it was evident that teachers in School A (T1 and T2) ignored the writing section of unit 2, while teachers in School B (T3 and T4) taught the writing section in unit 4. Teachers in School A (T1 and T2) explained that the writing tasks in unit 2 were not tested in the exams and were not worth the time from their perspectives. In this case, the curriculum was narrowed down to focusing mainly on what was tested rather than fostering language skills (Choi, 2008). The analysis also demonstrated that the Zhongkao writing task and other writing tasks of formal exams had tested students' ability to translate from Chinese to English, which encouraged teachers to focus on teaching translation as an alternative instead of writing. The task instructions were mainly written in Chinese, with a few English keywords attached, and required students to include the main points provided in their articles. In this way, the task-completion process involved students translating from the instructions and composing the article without their own planning. Such a phenomenon might result from the long-standing tradition of using the grammar-translation method in the context (Gan et al., 2018). As some teacher participants (T1 and T4) argued, such a design could support students in writing a complete article. For the markers, the design might also be convenient to them. However, a person could not learn how to ride a bike without taking off the training wheels. Similarly, the students' writing skills could not be developed without having them write on their own.

Besides the Zhongkao, the textbook mandated for this context also stood out as a major influence on teachers' teaching and assessment. The analysis showed that the design of the textbook attempted to strike a balance between teaching language knowledge and teaching communication but placed a visible emphasis on reading, vocabulary, and grammar. This emphasis was reflected in the teachers' teaching and assessment planning, as it was observed that the teachers spent relatively more time teaching and assessing these areas. Other aspects of language, such as listening, speaking, and writing, were included in the textbook design but received less space and suggested time in the teacher guidebook. The intention of the textbook design to highlight the importance of communication was represented but not appropriately enough. The teacher guidebook was also discovered to have received experienced teachers' endorsement for its suggestions on how they should plan teaching. Novice teachers, however, considered many activities subscribed in the guidebook not matching their students' language levels. It is difficult to blame the textbook and guidebook for such a circumstance, as most materials are designed for an idealised group of students

with comparable language proficiency and cannot manage to accommodate the different needs, learning strategies, and attitudes of individual learners by themselves (McDonough et al., 2013). The teachers would need to adapt the activities proposed by the guidebook to the students' language levels and needs instead of replacing them with other activities that did not fulfil the original aims.

With the Zhongkao powerfully affecting the context with its high stakes and the textbook defining what and how teachers teach, the impact of the curriculum and the formative assessment policy seemed quite feeble. The curriculum, which held a balanced philosophy of language teaching and assessment, has set the tone for the textbook to highlight the importance of teaching language forms and meanings and assessing them with diverse assessment approaches (Gu, 2012; Liu & Wang, 2020). However, a lack of alignment has revealed itself between the curriculum and the Zhongkao in two aspects. The first aspect is that while the Zhongkao claims itself to be both an achievement test, which is based on the curriculum, and a selection test, which aims to assign educational resources (Shenzhen Admission and Examination Office & Shenzhen Institute of Education Sciences, 2021), the selection purpose of Zhongkao prevails, making the Zhongkao as the guiding principle of the forms and purposes of teachers' classroom assessment. The second aspect relates to the test design. The Zhongkao placed a higher emphasis on assessing language knowledge and reading skills, driving teachers to spend excessive time teaching and assessing these areas instead of fostering students' communication skills as a whole, which is envisioned by the curriculum. The writing task design also demonstrated a lack of alignment since translation was not included as a skill that students should master in the curriculum. The Zhongkao has become the curriculum for the teachers and students, while the actual curriculum has become a regular paper document that receives little attention from teachers.

At the same time, the formative assessment policy, which was released two decades ago and has been reflected in the curriculum, has received scant attention in three teachers' (T1, T2, and T4) classrooms. The teachers seemed largely unaware of the formative assessment policy, what the policy promotes, and even the guidance on formative assessment provided by the curriculum (see section 1.3, pp. 12-13). Besides the many factors at the classroom and school levels, which have been discussed in the above sections, one of the reasons behind the teachers' ignorance of formative assessment might be that the teacher guidebook, which serves as the bridge between teachers and the curriculum, contained little detailed guidance

that instructed teachers to perform formative assessment activities. From the analysis, it can be noticed that the guidebook has introduced some teaching techniques and activities that had formative implications. However, the term ‘formative assessment’, or any other alternative term that refers to the assessment activities promoting learning in classroom, was not mentioned once in the guidebook. Exercises featuring closed tasks, which were widely adopted in summative exams, took up a much larger proportion of the assessment activities provided. Arguably, it was not possible for the guidebook to provide step-by-step advice for teachers’ formative assessment activities, as the validity of formative assessment is based on whether the feedback involved could effectively improve students’ learning, and for different students under different circumstances, the feedback process could vary significantly (Stobart, 2012). However, it cannot be an excuse for the absence of formative assessment in the teacher guidebook. To raise teachers’ awareness of implementing formative assessment, the guidebook could provide an introduction to the formative assessment concept and offer more explicit assessment instructions with example classroom assessment activities.

7.4 A framework for understanding English language teachers’ classroom assessment activities and assessment beliefs

Based on the discussion in the previous sections, I propose a framework for understanding English language teachers' classroom assessment activities and assessment beliefs in the Chinese middle school context (see figure 7-1). This framework addresses the relationship between elements within teachers’ classroom assessment activities, their understanding of learning, teaching, and assessment, and their past and present experiences and future aspirations in the situated context. It also includes factors identified in the teacher- and school-related practices from the dataset to highlight what teachers and schools consider as practical issues in assessment. To the best of my knowledge, no framework has been proposed to offer a comprehensive overview of teachers’ assessment activities, their beliefs behind their practices, and the local factors contributing to their actions and thoughts in the Chinese middle school context.

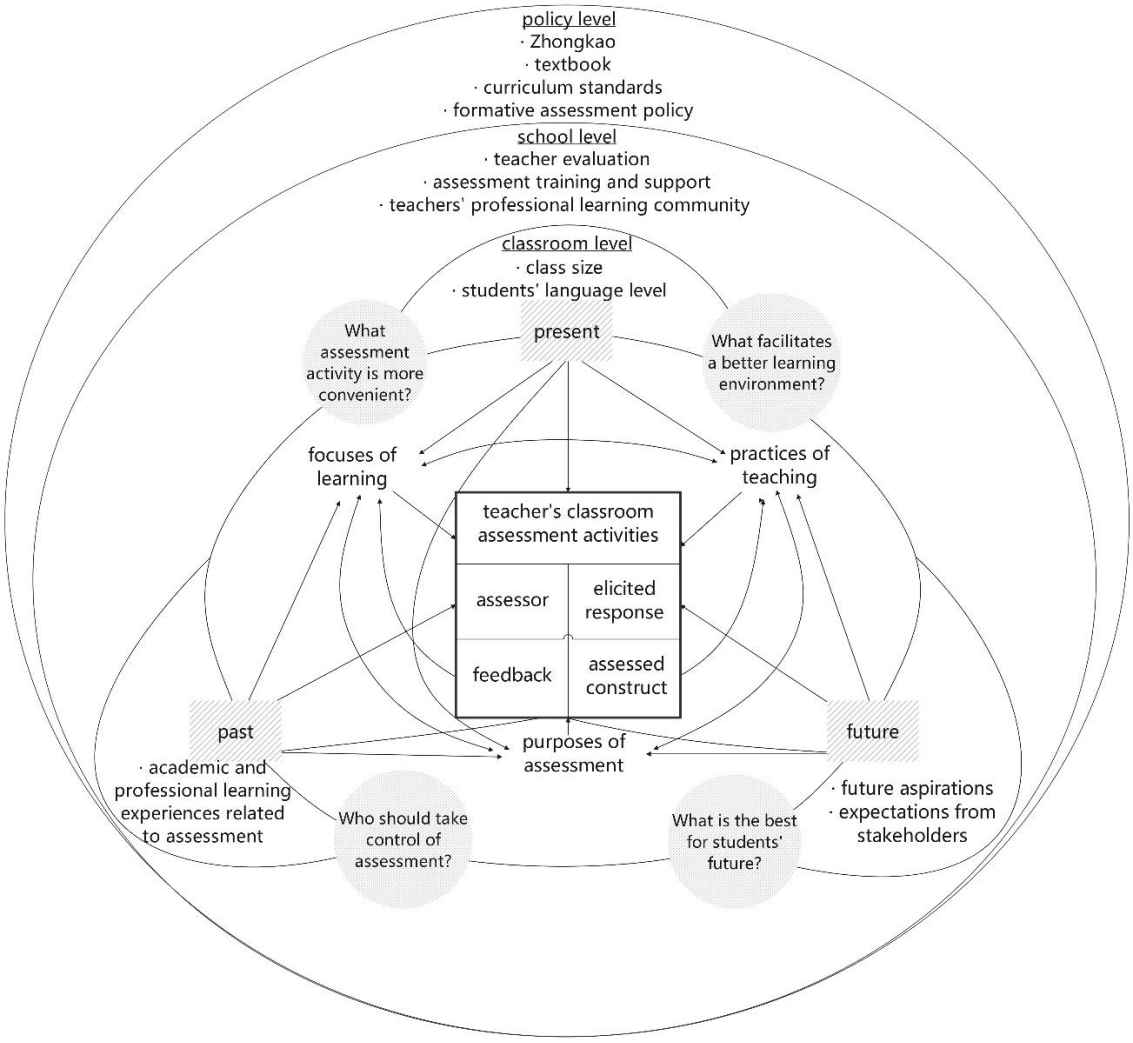


Figure 7-1 A framework for understanding English language teachers' classroom assessment activities and assessment beliefs in the Chinese middle school context

The square in the middle represents teachers' classroom assessment activities, including four elements that deserve close attention in analysis: assessor, elicited response, assessed construct, and feedback. The identification of the elements is inspired by Hill and McNamara's (2012) dimensions of classroom-based assessment, which highlight the approach to collecting assessment data (what is assessed and how the evidence is collected) and the target and agent of the assessment (who is assessed by whom). This framework further emphasises the role of feedback, which plays a central role in improving learning.

The four elements signify four sets of questions that should be answered while investigating teachers' classroom assessment activities. The first set of questions is about the assessors.

To understand the assessment activities, it is important to first identify who the assessors are, whether it is teacher or student. If students act as the assessors of their work or their peers' performances, it is worth discovering whether they assess autonomously or under the teacher's instruction. The second set of questions is about the elicited responses. Students' classroom responses can take different forms, oral or written, single or group. They can help researchers look at how the teachers generally plan their assessment activities in class. The third set of questions is about the assessed constructs. Constructs, the underlying ability that assessment intends to measure (Hughes, 2003), are the core of assessment activities and represent what teachers believe matters in learning, teaching, and assessment. Whether teachers focus on successfully completing a task or more open engagement with the task also requires specific consideration. The fourth set of questions is about the feedback. The quality of the feedback provided determines the success of teachers' formative assessment activities (Stobart, 2012). The kind of feedback received by students and whether such feedback focus on accuracy or learning promotion must be recognised.

The portraits of teachers' classroom assessment activities set up a foundation for understanding what teachers' understandings of assessment and formative assessment are. By analysing the assessment activities, three aspects of the teachers' beliefs should be understood: their focuses of learning, practices of teaching, and purposes of assessment. These aspects draw from Hill and McNamara's (2012) dimensions of classroom-based assessment, which emphasise the values guiding the teachers' assessment and how the evidence is used, and James's (2008) classification of different views of learning and assessment. The framework highlights the connections between learning, teaching, and assessment and argues that teachers' focuses of learning, practices of teaching, and purposes of assessment are often consistent to a certain extent within a teacher's mindset.

The focuses of learning refer to the knowledge and skills teachers consider as the key objectives. In the case of language teaching, it is essential to distinguish the teachers' beliefs about language and language learning. Lightbown and Spada's (2013) and Mitchell and Myles's (2004) reports on language theories are good points to start. Such beliefs are embedded in teachers' understanding of what language is and how language should be learnt, taught, and assessed. The practices of teaching refer to how teachers teach to pass the knowledge. Some teachers consider that teaching means transmitting knowledge for students to absorb; some believe that teaching needs to guide students to make sense of the knowledge

and to build a model to understand the world; some regard teaching as organising events that involve students in learning through actions (James, 2008). The purposes of assessment address how teachers view the function of assessment and what they incorporate assessment into their classrooms for. They might conduct assessment for judgments, learning, teaching, or classroom management. These three aspects of teachers' beliefs are often interconnected with each other and inform one another in actual practices.

After depicting teacher's assessment activities and beliefs, the focus is shifted to exploring the factors influencing the teachers' work. This framework classified the factors that influenced teacher's classroom assessment activities and beliefs into three categories: past, future, and present. The classification draws on Priestley and colleagues' (2015) teacher agency model, which explains teacher agency from the teacher's past experience, orientation towards the future, and engagement with the present. The framework also refers to Braun et al.'s (2011) four contextual dimensions framework for policy enactment (situated, professional, materials, and external contexts) to locate specific elements of teachers' work environment and life experiences that contribute to their assessment activities and beliefs. These perspectives compose a system in which teacher's assessment activities and beliefs are afforded and constrained.

To begin with, teacher's academic and professional learning experiences related to assessment are sought to portray teacher's experiences from the past, which contributes to understanding teacher's capacity and where teacher's beliefs originate. Then, factors linked to the future, including teacher's aspirations and stakeholders' expectations, should be investigated in terms of their effects on teacher's assessment practices and beliefs. At last, factors related to teacher's present assessment experiences are explored and grouped into three levels: classroom, school, and policy. Factors at the classroom level, such as class size and students' language level, directly influence teachers' assessment activities design. Some factors at the school and external policy levels, such as school's evaluation, assessment training and support for teachers, textbook, and test policy extend a more considerable impact on what teachers experience now, in the past, and in the future. Their effects stretch over a longer term, determining the type of professional learning experiences they gained, the present work environment they adapt to, and the aspirations they hold towards the future. Some factors at these two levels, including teacher learning community, curriculum standards, and formative assessment policy, however, show less extensive influence due to

the inadequate professional training teachers received and the strong impact of textbook and high-stake test design on teachers' teaching and assessment.

The framework also includes factors identified from the dataset to highlight what teachers and schools consider as practical issues in assessment. These factors are framed as questions raised from the teacher- and school-related practices and are rarely addressed by current assessment theories. The first question is 'what assessment activity is more convenient?' According to the data, a main reason for the teachers to choose teacher-led oral and written assessment was that these methods were convenient and efficient in assessing while teaching and in assessing large groups of students. For teachers who have heavy teaching workloads and are responsible for large classes, whether an assessment activity is convenient to implement could be a significant issue. The second question is 'what facilitates a better learning environment?' In the dataset, some teachers considered classroom discipline as a crucial factor that ensures a better learning environment and adopted assessment activities to make students perform in the same way and obey certain rules, while some teachers believed communication matters more in language classrooms and adopted more discussion activities without being worried about the 'messy' situation they might cause. The teachers' understanding of what contributes to the construction of a better learning environment, therefore, may impact how they design their classroom assessment in general.

The third question is 'who should take control of assessment?' From the interview data, it was discovered that teachers' beliefs about what role students should play in assessment and students' capacity to assess themselves could influence whether teachers hold a positive view towards and implement more student-assessed activities. Teachers who consider students as passive recipients of assessment who are not able to assess themselves accurately might adopt more teacher-controlled activities, with the teacher being the designer, the initiator, and the feedback provider, while teachers who recognise students' potential as assessors and create opportunities for them to enhance their assessment skills might undertake more activities involving students as more active roles. The final question is 'what is the best for students' future?' This involves both teachers' and schools' interpretation of what the educational goals are at the present stage of education and what they should do to achieve these goals. For the context of this study, exam results are something that the teachers and schools have to work for. However, there are other goals that also matter for the students' overall development, for example, their attitudes toward learning, their metacognitive

strategies, their psychological well-being, and their moral development. How the teachers and schools set their priorities for their goals may be reflected in how they plan their assessment activities and assessment-related policies.

7.5 Summary

This chapter discusses the findings in response to the three research questions of this study. The assessment activities implemented in the two schools and the teachers' different views on language learning and assessment have been discussed. The two teachers in School A (T1 and T2) held a rather behaviourist position in planning and viewing language teaching and assessment; one teacher (T3) in School B held a constructivist view of teaching and assessment, while the other (T4) held a mixed belief, combining constructivist objectives with behaviourist approaches. The purposes behind the teachers' assessment activities have been explored to explain teachers' assessment design and conceptions further. The teachers were found to assess for planning their instruction, classroom management, and enhancing students' learning (though many techniques were discovered not to promote learning significantly). The impact of multiple factors in the assessment context has also been discussed from the iterational, projective, and practical-evaluative dimensions. The iterational dimension witnessed teachers struggling with limited professional education opportunities regarding classroom assessment. The projective dimension saw teachers' future aspirations be fundamentally shaped by the high-stake selection test Zhongkao and the expectations of other educational stakeholders. The practical-evaluative dimension has made an influence on teachers' assessment activities and beliefs from the classroom, school, and external policy levels: local affordances and constraints, such as class size and students' language level, have formed teachers' practices at the classroom level; school evaluations, local support for teachers' assessment, and local teacher learning communities have shaped teachers activities and beliefs at the school level; the external policies, particularly the Zhongkao and the textbook, have become the curriculum that guided the teachers' work. The chapter concluded with a framework for understanding English language teachers' classroom assessment activities and assessment beliefs in the Chinese middle school context.

Chapter 8 Conclusion

In the previous chapter, I discuss the findings of this study from three perspectives: assessment activities in the two schools, teachers' purposes of assessment, and the assessment context. In this final chapter, I conclude by reflecting on the whole study from five aspects. Firstly, I address, in summary, the three research questions of this study (section 8.1). Secondly, the significance of this study is elaborated (section 8.2). Thirdly, the limitations of this study are summarised (section 8.3). Fourthly, the recommendations of this study are discussed in terms of assessment practices of different stakeholders (section 8.4). Finally, I address the future directions stemming from this study (section 8.5).

8.1 Addressing the research questions

In Chapter 1, three research questions were proposed to explore the research problem identified in this study. For the convenience of the readers, the three sets of questions are again listed below.

> RQ1: What classroom assessment activities do teachers adopt? Why do they choose these activities?
>
> RQ2: What are teachers' understandings of assessment, in particular formative assessment? Why do they adopt such beliefs?
>
> RQ3: How is the current assessment policy reflected in the textbook, teacher guidebook and the Zhongkao teachers work with? What actions have the teachers taken to put the assessment policy into practice under the influence of these materials? Why do they enact the assessment policy as such?

Regarding classroom assessment activities, the study identifies three types of assessment activities in the four teachers' classrooms: oral assessment activities, written assessment activities, and student-assessed activities. As reported in section 5.2.2 (p. 95), section 5.3.2 (p. 103), section 6.2.2 (p. 129), and section 6.3.2 (p. 138), in all classes, oral assessment activities were adopted the most often to check on students' mastery of textbook knowledge; written assessment activities fashioned after high-stake test tasks were implemented less

frequently; student-assessed activities were conducted the least often. It was also discovered that the teachers' feedback was mainly restricted to confirmation and correction at the task level, while only one teacher (T3) was observed to offer advice to students at both task and learning strategy levels.

From the analysis, the teachers mainly adopted assessment activities in English classrooms to evaluate whether students could respond to their questions correctly and offer students opportunities to drill for tests. They considered oral assessment activities efficient methods to implement (T1, T2, and T4) and to assist students' learning (T3) and regarded written assessment activities as manageable assessment methods that could be easily operated in large classes (T1, T2, T3, and T4). It was revealed that they planned their assessment activities for purposes other than learning, including assessing for organising their instructions and assessing for managing classrooms. The teachers' views diverged significantly regarding student-assessed activities. Teachers in School A (T1 and T2) showed less interest in conducting student-assessed classroom activities because of the lack of confidence in students' abilities to assess or their ability to implement such activities. On the other hand, teachers in School B (T3 and T4) displayed more positive attitudes towards student-assessed activities and were observed to engage students in assessment activities more frequently; however, the student-assessed activities in their classes only took up a minor proportion. In general, teacher-centred assessment and feedback at the task level gained prevalence among the teachers, while student-assessed activities and feedback on learning strategies and processes received inadequate consideration.

Regarding teachers' understandings of assessment, the study begins by analysing the teachers' beliefs of what matters in language learning and teaching and then moves on to identify where their assessment views sat. As shown in Chapter 5 (see p. 96/104), the two teachers in School A, T1 and T2, generally adopted closed questions and tasks to assess students' mastery of vocabulary, grammar, and reading skills, as they considered them the key objectives of their teaching. On the other side, as demonstrated in Chapter 6 (see pp. 127-128/136-137), T3 and T4 reported that they intended to foster and assess students' abilities to apply language knowledge in a situated context. However, only T3 was observed to implement more open questions and tasks to construct a context for students to perform. T4 held a seemingly contrasting belief, combining constructivist-oriented aims with behaviourist-oriented strategies. Although she voiced the hope for developing students'

communicative language skills, her actual practices were restricted to the grammar-translation approach, which prioritised memorisation and translation and hardly practised students' communication effectively. It seemed that behaviourist-oriented thinking and approaches, which define language as separate linguistic components and skills and language assessment as assessing whether students could reproduce the discrete knowledge they transmitted (Brown & Abeywickrama, 2019; James, 2008), were common among the teachers' practices. Constructivist-oriented communicative approaches, which highlight language use for meaningful communication (Brown & Lee, 2015), were not prevalent options in these classes.

The teachers' beliefs regarding formative assessment appeared to be consistent with their beliefs concerning language learning and assessment. T1, T2, and T4 reported that they regarded formative assessment as continuous assessments of students' performance or ongoing bite-size tests, which gave little attention to students' further learning improvement. They seemed to adopt a more behaviourist stance, in which formative assessment was interpreted from a teacher-directed perspective, and learning was based on reinforcement and correction (Carless, 2011). Such a stance resonated with their tendency to teach and assess language as discrete elements and apply more teacher-centred assessment activities. On the other hand, T3 viewed formative assessment as developing better learning strategies that promoted students' learning in the long run. She seemed to accept that students' self-direction matters in assessment, an idea grounded in the constructivist learning principles. Although her class did not exclude assessment activities with behaviourist characteristics, T3 has incorporated some core features of formative assessment, which involves enabling students to become self-regulated learners, in her practice and thinking.

Regarding teachers' understandings and enactment of the current assessment policies, the study addresses three levels of assessment policies, including the school policies, the Zhongkao policies, and the curriculum policies. As revealed in section 5.5.1 (see p. 117) and section 6.5.1 (see p. 152), for the school policies, most teachers agreed that they attempted to improve students' test performance through classroom assessment activities because the schools judged the quality of their teaching through students' test scores, while only one teacher (T3) reported that she was not affected by such policies, possibly due to her senior position in School B. The impact of teacher evaluation policies was thus revealed. For the Zhongkao policies, section 5.5.2 (see p. 118) and section 6.5.2 (see p. 153) displayed that

the Zhongkao has exerted complicated washback effects on teachers' practices. Despite the test designers' intention to encourage teachers to teach for communication, the teachers interpreted the focus of Zhongkao from a different perspective, the details of which will be elaborated in section 8.2.2. As for the curriculum policies, the teachers largely ignored the role curriculum played in their teaching and assessment. As demonstrated in section 5.5.3 (see p. 121) and section 6.5.3 (see p. 154), they hardly referred to the curriculum for instructions on their practices and primarily relied on textbooks and Zhongkao as references for teaching and assessment planning. What seemed to guide the teachers' teaching and assessment were the Zhongkao and the textbook they used.

The study also set out to discover the reasons behind teachers' assessment activities, their understandings of assessment and assessment policies, and their enactment in classrooms. In section 6.4.3 (see p. 150) and section 6.5 (see p. 152), several factors have been identified to have influenced how the teachers practised and understood classroom assessment and need to be reported as a system. First, the teachers were impacted by their past professional education. Having received insufficient input regarding assessment theories and approaches, the teachers were further affected by the inappropriately delivered assessment knowledge from the local teacher educators and the prevailing idea that teaching matters more than assessment in teachers' work. The lack of previous professional development in assessment has blocked teachers from referring to past experiences for assessment examples and led them to a situation where they need to constantly carry out trials with different methods. Second, the teachers were also driven by the local expectations that students should achieve higher scores to gain access to better educational resources in the future. Such expectations resulted from limited educational resources and have significantly shaped how teachers taught and assessed. Teaching to the tests became a routine, and improving students' test performance became the centre of the teachers' feedback. Third, factors within the teachers' present working environment were also observed to contribute to the teachers' actions and thinking. The large class size and students' English language level restricted teachers' choices of assessment at the classroom level; the accountability mechanism, the limited assessment learning opportunities the schools provided, and the local teacher learning communities afforded and constrained their practices and understandings at the school level; the Zhongkao and textbook design influenced teachers' conceptions of what needed to be taught and their classroom conducts at the policy level. Four questions raised from the data are also included in the framework to highlight what teachers and schools consider as

practical issues in assessment. Together, these factors constructed a framework in which teachers negotiated their assessment practices, understandings, and their enactment of the current assessment policies.

8.2 Significance of the study

The significance of this study is threefold. First, from a theoretical perspective, this study proposes a theoretical framework that seeks to explain Chinese middle school English teachers' classroom assessment activities and assessment beliefs by drawing on theories of multiple domains (section 8.2.1). The framework challenges the cultural assumptions in some formative assessment literature and offers a contextual standpoint to understand formative assessment more deeply in the Chinese middle school context. Second, from a local perspective, this study contributes to the literature in the Chinese middle school English education domain and deepens the local understanding of the Zhongkao reform in Shenzhen in terms of middle school English teachers' enactment of the local assessment policies and reactions to the new test (section 8.2.2). Third, from a personal learning perspective, this study allows me to go on a learning trip to reshape my understanding of classroom assessment in English classrooms in China (section 8.2.3). These perspectives are explained as follows.

8.2.1 Theoretical perspective

This study proposes a theoretical framework for understanding Chinese middle school English teachers' classroom assessment activities and assessment beliefs and contributes to the theoretical understanding of formative assessment by drawing on theories of multiple domains. Inspired by Hill and McNamara's (2012) dimensions of classroom-based assessment, the framework first identified key elements in teachers' classroom assessment activities, including assessor, elicited response, assessed construct, and feedback. It argues that, to achieve a comprehensive view of teachers' classroom assessment activities design, the target and agent of the assessment, the approach to collecting assessment data, the construct of the assessment task, and the role of feedback should be observed and understood. The framework also draws on James's (2008) classification of different views of learning and assessment and various strands of language learning theories, including behaviourist theory, communicative learning theory, and sociocultural learning theory, to explore the

connections between teachers' classroom assessment activities and assessment beliefs. It suggests that teachers often hold relatively consistent views when they choose the focuses of learning, plan their practices of teaching, and decide on their purposes of assessment. Their actions, however, are often under the influence of external factors from multiple aspects, making it necessary for the framework to further address these factors from both contextual level and teacher agency level.

The framework draws on Priestley and colleagues' (2015) teacher agency model to categorise different contextual factors and refers to Braun et al.'s (2011) contextual dimensions framework for policy enactment to identify elements contributing to the shaping of teachers' actions and beliefs. The factors are categorised based on whether they exert influences from the teachers' past, whether they represent or affect teachers' orientation towards the future, and whether they have an effect on teachers' engagement with the present. The influences from the past mainly include teachers' academic and professional learning experiences related to assessment. The orientation towards the future mainly involves the impact of high-stake selection test on teachers' views of assessment and the expectations of local educational stakeholders regarding test results. The engagement with the present concerns factors from classroom level, school level, and policy level. At the classroom level, class size and students' language level determine the range of assessment activities teachers can choose from; at the school level, the teacher evaluation policies, the available assessment training and professional support, and the assistance teachers can achieve in the local teacher learning community jointly form teachers' work environment and reshape their beliefs towards assessment; at the policy level, Zhongkao, the high-stake test impacting the local context, the mandated textbook, the curriculum standards, and the formative assessment policy exert influence on the local classrooms to various extent. The framework also raises four questions to reflect on teachers' and schools' consideration of the practical issues in assessment, which are rarely mentioned by current assessment theories.

Instead of viewing formative assessment as a practice shaped by the local culture, the framework understands formative assessment from a contextual perspective. In previous literature, China has been portrayed as one of the representatives of the Confucian-heritage culture contexts, in which summative assessment dominates the classroom settings, and student-centred pedagogical and assessment ideas find it challenging to become established. For instance, Brown et al. (2011) argue that differences in culture and society can lead to

distinctive assessment policies, practices, and conceptions. Carless (2011) also discusses formative assessment practices from a cultural perspective. Although he attempted to stand in the middle and avoid stereotyping particular contexts, his work was widely cited for supporting arguments that certain features belong to the Confucian-heritage culture contexts. Culture has been put forward as a taken-for-granted explanation for unsuccessful attempts and challenges in formative assessment practices in these contexts. However, such an explanation can be overused in understanding local assessment practices, of which the reasons are listed as follows.

In the first place, many phenomena described as the features of the Confucian-heritage culture, such as the strong faith in summative tests, teaching to the tests, and the utilitarian values of education, are not exclusively owned by the contexts with the Confucian-heritage label. Studies undertaken in countries around the globe have identified these phenomena in classrooms with distinct cultures (for example, Ali & Hamid, 2020; DeLuca et al., 2021; Mansell et al., 2009). My personal experience of interacting with international audiences in conference presentations also corroborated this account. For example, when I presented my data about the teachers' adopted assessment methods and their conceptions of those methods in the Language Testing Research Colloquium, one of my audience from Mexico approached and said, "lots of the findings can relate to the Mexican context." It is clear that Mexico enjoys a distinct culture from China and has not been labelled as a Confucian-heritage context. Therefore, this study suggests that ascribing the challenges in formative assessment implementation to cultures or the influence of Confucianism should be considered critically, and some challenges might be universal instead of unique to certain contexts.

Secondly, the learning culture of a context and its embedded values can be dynamic and changeable due to the fast-changing local circumstances. Studies done in the Chinese context, for example, have recounted such changes. For instance, the Shanghai students in Shi (2006) hold similar beliefs about active learning and interactive teacher-student relationship with their western counterparts, and students' needs, efforts, and self-reflection are also highly valued by teachers in English as a foreign language (EFL) classrooms in a major city in East China (Xiao & Yang, 2019). Following these changes in learning culture, assessment beliefs of policymakers, teachers, and students have also made corresponding changes. The formative assessment initiative proposed by the Ministry of Education two decades ago demonstrated the determination of policymakers at the top to transform classroom

assessment; in-service and pre-service teachers were found to acquire a more comprehensive understanding of effective assessment (Coombs et al., 2021; Xu & He, 2019); students were reported to respond more positively to assessment aiming at advancing their learning, skills and moral characters (Chen & Brown, 2018). Therefore, the 'teacher-centred tradition' and the 'testing tradition' in China are not as entrenched as some researchers claimed, and it would not be rigorous to label the entire context with cultural stereotypes and findings in the past.

Truly, referring to other contexts and identifying the differences between one's practices and others' could be easily accessible ways of reflecting on oneself – to be honest, I took the same stance at the beginning of my PhD journey. However, overusing culture as an explanatory framework can lead to the oversimplification of the problems teachers face and the negligence of other factors that impact teachers' assessment practices and understanding. Then, what might explain the challenges formative assessment initiatives encountered in local classrooms? This study argues that the local factors influencing teachers' practices and beliefs from different perspectives should be investigated to uncover the local constraints. In this study, the teachers experienced inadequate learning experiences regarding assessment, high expectations of students' test performance, large classes with low-proficiency students, restricted support from schools, and inappropriately represented curriculum ideas in the Zhongkao and textbook. These contextual factors are the results of the ineffective teacher education mechanism, the limited educational resources, the insufficient recognition of the importance of formative assessment, and the drawbacks in the high-stake test designs and teaching materials. They cannot be explained as cultural phenomena.

For other contexts where formative assessment implementation is also confronted with difficulties, it is essential for the local practitioners and researchers to closely observe the context they work in and ponder over the connection between different influences on teachers' practices. The framework for understanding classroom assessment activities and teachers' assessment beliefs proposed in this study can serve as a reference for other contexts, but adaptations will need to be made according to the local situation. Black and Wiliam (2018) argue that formative assessment needs to be understood within the situated locality. To comprehend local formative assessment practices and understandings, exploring contextual factors matters more than overplaying the influence of culture.

8.2.2 Local perspective

This study enriches the literature on Chinese middle school English teachers' assessment activities and understandings. While Chinese high school English teachers have received much attention regarding their assessment practices and beliefs from Chinese and overseas academics due to the influential high-stake assessment Gaokao, middle school English teachers who strive for the Zhongkao, another equally important test at their stage, receive limited academic attention. As a critical period at the end of compulsory education, middle school education builds a foundation for students to develop their knowledge and thinking skills for further education. As for English education, it is crucial for students to acquire an international language during this period to enhance their abilities to communicate across cultures, become reflective and self-managing, and learn other knowledge through English (Ministry of Education, 2011). Whether assessment at this stage contributes to the development of students' communicative and learning abilities needs to be examined. If these aims have not been met, more efforts should be invested in understanding the existing problems and challenges and the factors that give rise to them.

The study also connects findings from assessment research done in the Chinese tertiary English education area with the phenomenon that took place in the middle school context. The term 'process assessment', adopted by College English teachers as an alternative for formative assessment (Chen et al., 2013), was identified as a widely accepted term by middle school English teachers. While College English teachers were found to interpret 'process assessment' as continuous testing (Chen et al., 2013), middle school English teachers were observed to have made different interpretations of the term, varying from frequent exams to fostering students' learning strategies. This adds to the assessment literature from Chinese middle school English teachers' perspective and reveals a complex picture in which the teachers negotiated the meaning of formative assessment and enacted what they believed was the best for the students' benefits.

Moreover, the study is very timely, as it was conducted at the start of the Zhongkao reform in Shenzhen when the teachers adjusted their teaching and assessment according to the new local high-stake assessment. The study revealed that the new English Zhongkao design had both positive and negative influences on teachers' practices and understandings. On the positive side, the new English Zhongkao has increased the proportion of the listening-and-

speaking test, encouraging teachers to spend more time practising listening and speaking in class. Even though the teachers were observed to merely prescribe more listening-and-speaking exercises, increasing scores for listening and speaking remained a good start of the assessment reform, as students were given more opportunities to practise their listening and speaking. The new English Zhongkao also allocated ample spaces for reading comprehension tasks and cloze tasks for vocabulary and grammar, ensuring that teachers devoted sufficient efforts to consolidating students' language knowledge and reading skills. However, on the negative side, the new English Zhongkao assigned only a small proportion of the total scores to listening, speaking, and writing. Such a disproportion in test design encouraged teachers to spend excessive time on teaching language knowledge and reading skills but overlooked other language skills in their teaching design. The overuse of Chinese instructions in writing tasks also led teachers to teach writing as teaching translation, which could adversely affect students' development of writing skills. These findings illustrate how the local middle school English teachers understood the new assessment changes and reacted to the changes with their classroom assessment activities, which have implications for future local test design. Evaluations of the washback effects on teachers' practices may enable more beneficial changes to make the local assessment reform successful.

8.2.3 Personal learning perspective

At the very beginning of this PhD journey, I, as a beginner in the academic world, identified my research problem and made assumptions about what I would discover based on my experiences as a student, a language teacher, and a curriculum designer. When I was a student, the English classes I joined were mainly led by teachers. Classroom assessment at that time featured teachers raising questions and assigning quizzes for students to respond. Test drilling was common, particularly at the end of a semester before formal examinations. After graduating and taking different roles in the education sector, I noticed that novel pedagogical approaches, such as the communicative approach, were introduced to many classrooms. However, student-centred activities remained uncommon in large classes, and many teachers regarded assessment as tests and lacked the awareness that classroom assessment could significantly improve students' learning. Therefore, I expected that the data generation might reveal many teacher-centred assessment activities and test-oriented assessment conceptions.

The findings of this study matched some of my experiences and guesses before data generation. For instance, the teachers' excessive use of teacher-led oral assessment activities, in particular, the IRE/F pattern (teacher's initiation – students' responses – teacher's evaluation/feedback) (Leung, 2004), for assessing students' understanding of textbook knowledge was within my expectation, as classroom conversations as such could take place easily without sophisticated techniques and careful planning. The inadequate implementation of student-assessed activities and some teachers' beliefs that formative assessment meant bite-size tests also matched my prediction, as the context under investigation is known for its exam pressure. As a result, teachers might emphasise teacher-controlled summative tests more than student-centred classroom assessment. Some findings could also be anticipated because of previous reports. For example, behaviourist and constructivist approaches and beliefs of language learning, teaching, and assessment were both identified in the observed classrooms and the teachers' interviews, which corresponded with other language education research conducted in similar contexts (for example, Liu & Wang, 2020; Zhang & Liu, 2014). Contextual factors at different levels, such as the large size of the teachers' classes, the school's evaluation policies for teachers, and the heavy influence of high-stake summative tests on local expectations of students' test performance and teachers' future aspirations, were also recognised by other researchers (for example, Yan, 2012, 2015; Yan et al., 2018).

Some findings, however, deeply challenged my expectations and were largely absent from current literature. To begin with, it was found that the local teacher educators did not successfully introduce the meaning and importance of formative assessment to the local middle school English teachers. Some teacher educators might have even been perceived to adopt a narrow interpretation of formative assessment themselves, associating the term with frequent testing throughout the teaching process. This might at least partly explain the prevalence of the term 'process assessment' in the Chinese context. As explained in earlier chapters (see section 5.4.2, p. 112, and section 6.4.2, p. 147), process assessment was adopted by many Chinese teachers to replace the term formative assessment, as they considered formative assessment as continuous tests throughout a student's learning process. Two teachers (T1 and T4) in this study reported that they learnt about process assessment through teacher educators and senior teachers, which suggested that such a restricted interpretation of formative assessment has been accepted by those who had the power to educate teachers with less experience. This might be a hidden reason for the test-oriented

assessment conceptions held by many teachers and the unsuccessful assessment policy enactment in this context.

The study also highlighted that even though many teacher participants viewed formative assessment, or 'process assessment' as they described it, as continuous tests, others had more sophisticated understandings of the concept. For example, one teacher (T3) regarded the term as representing the development of students' learning strategies and was observed to incorporate feedback about self-regulation in class to enable students' ongoing effective learning. Such an understanding of and practice in formative assessment corresponds with its core, which refers to enabling students to achieve their best (Assessment Reform Group, 2002). This suggests that some teachers in this context have realised the principles and significance of formative assessment and endeavored to integrate the concept into their classrooms.

8.3 Limitations of the study

This study is not without limitations. The dataset of this study involves four teachers in two middle schools, which is a relatively small one. Although the findings produced from the dataset might not be able to represent all teachers in the selected context, the study was designed to ensure the depth of understanding teachers' assessment activities and beliefs and the production of a framework that explains the reasons behind their actions and thoughts. The framework summarises the contextual factors which afforded and constrained the teachers' work, clarifies the relationship between these factors, and offers guiding principles for future research into teachers' classroom assessment practices. Such findings can only be achieved through "investigat[ing] a contemporary phenomenon in depth and within its real-world context" and can hardly be attained through large-scale surveys, in which the actual practices and underlying ideas of teachers are difficult to be observed and elicited (Yin, 2018, p. 15).

Another limitation might be that the teachers in the two schools were observed to teach different units in the textbook. A potential risk for such a plan was that the different topics and design of the units might have impacted how teachers planned their teaching and assessment. For instance, the topics of some units might be more intriguing, and teachers might find it easier to arrange student-centred activities in teaching these units. However,

the qualitative nature of this study determined that accurately representing the participants' social actions and consistently applying the methodological and analytic approaches matters more than the consistency of the participants' actions. Therefore, several strategies were adopted to ensure the validity and reliability of this study. Firstly, the data were generated with careful ethical considerations, including anonymising the teachers, sharing transcripts with the teachers, and positioning myself as a learner, to minimise the effect of the researcher on teachers' practices. Secondly, the data were constantly re-examined, compared, and synthesised so that the themes, sub-themes, and codes produced consistently and strongly supported the arguments. Thirdly, the observation schemes and transcripts were repeatedly and thoroughly checked to ensure that the definitions of the themes, sub-themes, and codes were not shifted. Fourthly, continuous comparisons were made within a case and between cases throughout the data analysis to make sure that the inferences made were sound.

8.4 Recommendations of the study

Based on the findings of this study, I identify several recommendations for teachers, school leaders, teacher educators, test designers, policymakers, and textbook designers.

Recommendations for teachers are generated from both assessment activity and assessment understanding perspectives. For assessment activities, teachers can incorporate more open questions and tasks in their classes to elicit students' thoughts and encourage students to come up with more than one answer to increase the opportunities to practise communication. Students should also be invited to participate in assessment activities that involve them as the initiators of assessment. For assessment understanding, teachers should develop a communicative-oriented view of language learning and assessment. Vocabulary, grammar, and reading skills are necessary for test-taking and accuracy in students' performance. However, other language skills, such as speaking, listening, and writing, require more attention so that students can be one step closer to becoming competent communicators in English. More extended feedback on learning strategies and self-regulation skills should also be included in teacher-student interactions so that students are given more than surface corrective feedback that might not be generalised to other tasks.

Recommendations are also generated for school leaders. School leaders should enrich their understandings of formative assessment and become aware of how formative assessment

may improve students' attainment. It is crucial for them to have more patience with teachers and create a less stressful workplace for them to try innovative teaching and assessment methods with no worries. Anxieties about test performance and rankings should be avoided. School leaders should also provide teachers opportunities to learn about new assessment theories and approaches. This might include inviting assessment experts to schools to give lectures and workshops to teachers, inviting skilled teachers from other schools who have successfully implemented formative assessment in classes to share their experiences, and welcoming these professionals to offer detailed suggestions on how teachers in their schools might improve in their assessment design and practices. Teacher learning communities within schools should also be promoted. School policies can be established to encourage teachers with different professional and subject backgrounds to learn from each other through classroom observations and discussions.

For teacher educators, the recommendations are related to how they should communicate the concepts of formative assessment and the curriculum ideas to teachers and school leaders. The analysis shows that the local teacher educators might not have developed a comprehensive understanding of formative assessment and have passed down restrictive interpretations of formative assessment to teachers. Therefore, the teacher educators must update their assessment knowledge and renew their course materials. They could also make amendments to how assessment knowledge is delivered. For example, they could arrange workshops involving teachers exchanging ideas and working collaboratively, enabling teachers to learn by doing. Another critical move is that teacher educators should inform teachers about their roles in liaising between curriculum and assessment policies and classroom learning. If the teachers cannot realise the initiatives advocated by the policies at the top, such as developing students as language communicators and implementing diverse assessment approaches to promote learning, it is unlikely that they will recognise the part they play in the bigger picture or undertake the pioneering initiatives.

Recommendations also arise for local test designers, since the Zhongkao tremendously impacts how teachers teach and assess in this context. Through materials analysis, it is discovered that the Zhongkao has attempted to direct teachers' attention to teaching communication by increasing the proportion of the listening-and-speaking test scores. This is a trend that needs to be encouraged. In the future, when teachers and students become more familiarised with the current test design, and the classrooms witness communicative

approaches being more widely accepted, the scores of the listening-and-speaking test can be further increased. The reading comprehension and cloze tasks, which constitute the largest share in the Zhongkao design, could be reduced to a moderate level, allowing listening-and-speaking test and writing tasks to achieve more attention from teachers and students. It is not to say that reading, vocabulary, and grammar are not essential in language learning and assessment. However, from the findings of this study, the disproportionate share of the total score for reading and cloze tasks has been counterproductive to students' language learning in the way that knowledge related to these tasks was given excessive consideration. For writing tasks, English instructions should be used instead of Chinese to avoid writing being considered the same as translation and vice versa.

The study carries implications for curriculum designers and textbook designers. While the current curriculum standards involve guidelines and suggestions for teachers' classroom practices, the teachers did not consider them helpful, as most of them were too general and difficult to implement in classrooms. Therefore, more specific suggestions need to be made with detailed classroom examples so that teachers can understand the intentions behind the design and imitate them in their practices. Textbook designers should bear the idea of formative assessment in mind while designing the textbook, incorporate more classroom activities that enable opportunities for formative assessment, and integrate more explicit formative assessment suggestions in teacher guidebooks. Other implications might be beyond one or two policymakers' control, such as the necessity for increasing local educational resources, reducing class size, and increasing the quality of vocational education so that students have more attractive options for further education. These require long-term local and even national efforts and might not be accomplished soon, but they are things that are worth striving for.

The study also offers evidence of how better to support the national aspirations for assessment and to inform future directions for policy planning. It is now two decades since formative assessment was first introduced to the Chinese education system. While some positive changes in classroom assessment have been witnessed in this study and in other studies, there remain many language teachers who are unclear about its significance and how ideas of formative assessment might be made real in practice. The national assessment policy should now take a step further and move beyond simply advocating diverse classroom assessment approaches to working with teachers to promote teacher education in formative

assessment. Researchers and practitioners who specialise in formative assessment should be involved in language curriculum design, teacher education programmes, and collaborative projects with local teachers. This combination of activities will increase awareness amongst teachers of formative assessment and will ensure the development of formative assessment approaches more adaptable to Chinese classrooms and foreign language learning of Chinese students.

8.5 Future directions

The study has offered reports of four Chinese middle school English teachers' assessment activities, assessment understanding, and enactment of assessment policies. It revealed the challenges teachers encountered in classroom assessment, the misconceptions that emerged in their thinking, and the contextual factors that impacted their actions and beliefs. The findings of this study have forwarded thinking from multiple aspects. From a theoretical perspective, the study has proposed a theoretical framework that understand localised formative assessment practices from a contextual perspective and has challenged the cultural assumptions adopted by much formative assessment literature. From a local perspective, the study has enriched assessment literature concerning Chinese middle school English teachers and offered timely suggestions for the ongoing local assessment reform. From a personal perspective, I have gained an updated understanding of the current situation regarding assessment practices in English classrooms at Chinese middle schools.

There remain many questions stemming from this study that require future research. First, the school leader's perspective should be added while researching teachers' actions and thoughts. As an important role in establishing school policies, the school leaders' understanding of assessment could significantly shape teachers' classroom assessment practices and conceptions. Given such an influence, it is worth noting whether and how school leaders' assessment conceptions, experiences, and practices shape the teachers' movements and views. Second, students' perspectives should be included in future research. While the teachers' activities and understandings in this study have been explored, the effects of their practices on students' learning and performance were unknown. What is more, the data revealed some students' actions in class, which could not be explained by the data at hand, for instance, why some students assessed their peers' answers without teachers' instructions, and why such phenomena were only observed in the two novice teachers' class?

Without students' viewpoints and reactions to teachers' assessment activities, a complete picture of the classrooms cannot be obtained.

Third, both schools in this study adopted a mentoring system, in which experienced teachers serve as mentors of early-career teachers to pass their professional and personal experiences and techniques. The study also involved a pair of mentor and mentee participants (T3 and T4 in School B), which generated interesting data demonstrating the similarities and distinctions between the two teachers' practices and understandings. However, it is unclear in the data how T4 as a novice teacher understood T3's approaches and beliefs and whether T4 had attempted to incorporate T3's ideas in her teaching and assessment planning. How the system works and what the mentor-mentee relationship brings to the early-career teachers in terms of educational practices and understandings require further research. Finally, the use of teachers' and students' first language in the classroom assessment activities may require more attention, as students' linguistic repertoire can be useful in expanding their English learning (Schissel, Morales, et al., 2019). With meaning-making and communication being the centre of the curriculum and the bilingual approaches to language teaching in this context, how the assessed constructs should be defined and re-defined and how the assessment approaches should be modified in classroom assessment activities to capture what is accepted in terms of the local notions of bilingualism deserve more academic spotlight (Schissel, Leung, et al., 2019).

With all these questions unanswered, this thesis is not the end of the story. When my PhD journey was coming to an end, the Ministry of Education issued an updated English curriculum 'English curriculum standards for compulsory education (2022)' (义务教育英语课程标准 2022 年版) in 2022. The document has reiterated the importance of aligning teaching, learning, and assessment, highlighted the role assessment plays in learning, and restated the requirement of balancing formative and summative assessment in classrooms (Ministry of Education, 2022). It has also made a heartening change in the assessment guideline section by including a case with a transcription of classroom conversation illustrating how teachers can design classroom assessment to discover what students can do and provide constructive feedback to scaffold students' thinking, which resonates with the suggestion made by this thesis. A drawback is that only one case was included in the guidelines, which was far from enough. This shows that the policy intention to emphasise the for-learning function of assessment remains unchanged. In the future, we may witness

whether and how educational policies, including the new curriculum document and the new Zhongkao reform, influence the teachers' actual assessment practices. Attention needs to be given to those identified contextual factors which impacted the teachers' work to see whether changes have been made and how changes affect teachers' work.

Reference

Adamson, B. (2001). English with Chinese characteristics: China's new curriculum. *Asia Pacific Journal of Education, 21*(2), 19-33. https://doi.org/10.1080/02188791.2001.10600192

Ali, M. M., & Hamid, M. O. (2020). Teaching English to the test: Why does negative washback exist within secondary education in Bangladesh? *Language Assessment Quarterly, 17*(2), 129–146. https://doi.org/10.1080/15434303.2020.1717495

Aljaafreh, A., & Lantolf, J. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *The Modern Language Journal, 78*(4), 465-483. https://doi.org/10.1111/j.1540-4781.1994.tb02064.x

Anderson, J. (1993). Is a communicative approach practical for teaching English in China? Pros and cons. *System, 21*(4), 471-480. https://doi.org/10.1016/0346-251X(93)90058-O

Arens, A. K., Morin, A. J. S., & Watermann, R. (2015). Relations between classroom disciplinary problems and student motivation: Achievement as a potential mediator? *Learning and Instruction, 39*, 184-193. https://doi.org/10.1016/j.learninstruc.2015.07.001

Arnfast, J. S., Jørgensen, J. N., & Holmen, A. (2010). Second language learning. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (3rd ed., pp. 419-425). Elsevier.

Assessment Reform Group. (2002). *Assessment for learning: 10 principles*. University of Cambridge.

Baird, J., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: Fields apart? *Assessment in Education: Principles, Policy & Practice, 24*(3), 317-350.

Bakker, J. I. (2012). Interpretivism. In A. J. Mills, G. Durepos, & E. Wiebe (Eds.), *Encyclopedia of case study research* (pp. 487-493). SAGE.

Ball, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy, 18*(2), 215–228. https://doi.org/10.1080/0268093022000043065

Ball, S. J., Maguire, M., Braun, A., & Hoskins, K. (2011). Policy actors: Doing policy work in schools. *Discourse: Studies in the Cultural Politics of Education, 32*(4), 625-639. https://doi.org/10.1080/01596306.2011.601565

Barlow, C. A. (2010). Interviews. In A. J. Mills, G. Durepos, & E. Wiebe (Eds.), *Encyclopedia of case study research* (pp. 496-499). SAGE.

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

BERA. (2018). *Ethical guidelines for educational research*. https://www.bera.ac.uk/publication/ethical-guidelines-for-educational-research-2018-online

Berry, R. (2011a). Assessment trends in Hong Kong: Seeking to establish formative assessment in an examination culture. *Assessment in Education: Principles, Policy & Practice, 18*(2), 199-211. https://doi.org/10.1080/0969594X.2010.527701

Berry, R. (2011b). Educational assessment in mainland China, Hong Kong and Taiwan. In R. Berry & B. Adamson (Eds.), *Assessment reform in education: Policy and practice* (pp. 49-61). Springer.

Biesta, G. (2010). Pragmatism and the philosophical foundations of mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social & behavioural research* (pp. 95–117). SAGE.

Biesta, G., & Tedder, M. (2007). Agency and learning in the lifecourse: Towards an ecological perspective. *Studies in the Education of Adults, 39*(2), 132-149. https://doi.org/10.1080/02660830.2007.11661545

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*(3), 347-364. https://doi.org/10.1007/BF00138871

Biggs, J. (2012). What the student does: Teaching for enhanced learning. *Higher Education Research & Development, 31*(1), 39-55. https://doi.org/10.1080/07294360.2012.642839

Black, P. (2001). Dreams, strategies and systems: Portraits of assessment past, present and future. *Assessment in Education: Principles, Policy & Practice, 8*(1), 65-85. https://doi.org/10.1080/09695940120033261

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*, 7-74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (1998b). *Inside the black box: Raising standards through classroom assessment*. King's College London.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*, 5-31. https://doi.org/10.1007/s11092-008-9068-5

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

Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. In R. W. Tyler (Ed.), *Educational evaluation: New roles, new means* (pp. 26–50). University of Chicago Press.

Borg, S. (2018). Teachers' beliefs and classroom practices. In P. Garrett & J. M. Cots (Eds.), *The Routledge Handbook of Language Awareness* (pp. 75-91). Routledge.

Bowen, G. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal, 9*(2), 27–40. https://doi.org/10.3316/QRJ0902027

Braun, A., Ball, S. J., Maguire, M., & Hoskins, K. (2011). Taking context seriously: Towards explaining policy enactments in the secondary school. *Discourse: Studies in the Cultural Politics of Education, 32*(4), 585-596. https://doi.org/10.1080/01596306.2011.601555

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. SAGE.

Braun, V., Clarke, V., Hayfield, N., & Terry, G. (2019). Thematic analysis. In P. Liamputtong (Ed.), *Handbook of research methods in health social sciences* (pp. 843-860). Springer.

Brooks, N. (1966). Language learning: The new approach. *The Phi Delta Kappan, 47*(7), 357-359. https://www.jstor.org/stable/20371592

Brooks, R., Te Riele, K., & Maguire, M. (2014). *Ethics and education research*. SAGE.

Brown, G. T. L., & Gao, L. (2015). Chinese teachers' conceptions of assessment for and of learning: Six competing and complementary purposes. *Cogent Education 2*(1), 1-19. https://doi.org/10.1080/2331186X.2014.993836

Brown, G. T. L., Hui, S. K. F., Yu, F. W. M., & Kennedy, K. J. (2011). Teachers' conceptions of assessment in Chinese contexts: A tripartite model of accountability, improvement, and irrelevance. *International Journal of Educational Research, 50*, 307-320. https://doi.org/10.1016/j.ijer.2011.10.003

Brown, H. D. (2000). *Teaching by principles: An interactive approach to language pedagogy*. Pearson Education.

Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). Pearson Education.

Brown, H. D., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practices* (3rd ed.). Pearson Education.

Brown, H. D., & Lee, H. (2015). *Teaching by principles: An interactive approach to language pedagogy* (5th ed.). Pearson Education.

Bryman, A. (2016). *Social research methods* (5th ed.). Oxford University Press.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*, 1-47. https://doi.org/10.1093/applin/I.1.1

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. (2011). *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: Potentials and challenges. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning* (pp. 167-182). Springer.

Chen, J., & Brown, G. T. L. (2018). Chinese secondary school students' conceptions of assessment and achievement emotions: endorsed purposes lead to positive and negative feelings. *Asia Pacific Journal of Education, 38*(1), 91-109. https://doi.org/10.1080/02188791.2018.1423951

Chen, Q. (2017). Localized representation of formative assessment in China: A regional study from a sociocultural perspective. *Frontiers of Education in China, 12*(1), 75-97. https://doi.org/10.3868/s110-006-017-0005-9

Chen, Q., Kettle, M. A., Klenowski, V., & May, L. (2013). Interpretations of formative assessment in the teaching of English at two Chinese universities: A sociocultural perspective. *Assessment & Evaluation in Higher Education, 38*(7), 831-846. https://doi.org/10.1080/02602938.2012.726963

Chen, Q., May, L., Klenowski, V., & Kettle, M. A. (2014). The enactment of formative assessment in English language classrooms in two Chinese universities: Teacher and student responses. *Assessment in Education: Principles, Policy & Practice, 21*(3), 271-285. https://doi.org/10.1080/0969594X.2013.790308

Cheng, H., & Ding, Q. (2021). Examining the behavioral features of Chinese teachers and students in the learner-centered instruction *European Journal of Psychology of Education, 36*, 169-186. https://doi.org/10.1007/s10212-020-00469-2

Cheng, L., & Curtis, A. (2010). The realities of English language assessment and the Chinese learner in China and beyond. In L. Cheng & A. Curtis (Eds.), *English language assessment and the Chinese learner* (pp. 3-12). Routledge.

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

Choi, I. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing, 25*(1), 39–62. https://doi.org/10.1177/0265532207083744

Chomsky, N. (1959). A review of B. F. Skinner's verbal behavior. *Language, 35*(1), 26-58. https://doi.org/10.2307/411334

Chomsky, N. (1981). *Lectures on government and binding*. Foris.

Clark-Gareca, B., & Gui, M. (2019). Chinese and American EFL teachers' beliefs about curricular and pedagogical practices: cross-cultural similarities and differences. *Language and Intercultural Communication, 19*(2), 137-151. https://doi.org/10.1080/14708477.2018.1456546

Coombs, A., Rickey, N., DeLuca, C., & Liu, S. (2021). Chinese teachers' approaches to classroom assessment. *Educational Research for Policy and Practice, 21*, 1-18. https://doi.org/10.1007/s10671-020-09289-z

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative & mixed methods approaches* (5th ed.). SAGE.

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice, 39*(3), 124-130.

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. SAGE.

Cumming, A. (2009). Language assessment in education: Tests, curricula and teaching. *Annual Review of Applied Linguistics, 29*, 90-100. https://doi.org/10.1017/S0267190509090084

Davies, A. (2014). Fifty years of language assessment. In A. J. Kunnan (Ed.), *The companion to language assessment* (Vol. 1, pp. 3-22). John Wiley & Sons.

Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing, 21*, 305–334.

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly, 43*, 393-415. https://doi.org/10.1002/j.1545-7249.2009.tb00242.x

DeLuca, C., Rickey, N., & Coombs, A. (2021). Exploring assessment across cultures: Teachers' approaches to assessment in the U.S., China, and Canada. *Cogent Education, 8*(1), 1-26. https://doi.org/10.1080/2331186X.2021.1921903

Deng, C., & Carless, D. (2010). Examination preparation or effective teaching: Conflicting priorities in the implementation of a pedagogic innovation. *Language Assessment Quarterly, 7*, 285–302. https://doi.org/10.1080/15434303.2010.510899

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2005). *The SAGE handbook of qualitative research* (3rd ed.). SAGE.

Emirbayer, M., & Mische, A. (1998). What is agency? *American Journal of Sociology, 103*, 962–1023. https://doi.org/10.1086/231294

Fan, X. (2006). 回首百年话科举——论科举制度历经的沧桑及对现代高考制度的影响 [Contemplation on the imperial examination system]. *北京工业大学学报(社会科学版) [Journal of Beijing University of Technology (Social Sciences Edition)], 6*(1), 92-95 (in Chinese).

Feng, Y. (1999). National college entrance examinations: The dynamics of political centralism in China's elite education. *The Journal of Education 181*(1), 39-57.

Fleer, M. (2015). Developing an assessment pedagogy: The tensions and struggles in re-theorising assessment from a cultural–historical perspective. *Assessment in Education: Principles, Policy & Practice, 22*(2), 224-246. https://doi.org/10.1080/0969594X.2015.1015403

Forsberg, E., & Wermke, W. (2012). Knowledge sources and autonomy: German and Swedish teachers' continuing professional development of assessment knowledge. *Professional Development in Education, 38*(5), 741-758. https://doi.org/10.1080/19415257.2012.694369

Fu, K. (1986). *中国外语教育史 [A history of foreign language teaching in China]*. Shanghai Foreign Language Education Press.

Fulcher, G. (2020). Operationalizing language assessment literacy. In D. Tsagari (Ed.), *Language assessment literacy: From theory to practice* (pp. 8–28). Cambridge Scholars.

Gan, Z., Ieong, S. S. L., Su, Y., & He, J. (2018). Understanding Chinese EFL teachers' conceptions and practices of assessment: Implications for teacher assessment literacy development. *Australian review of applied linguistics, 41*(1), 4-27. https://doi.org/10.1075/aral.17077.gan

Gardner, J. (Ed.). (2012). *Assessment and learning* (2nd ed.). SAGE.

Gibbs, G. R. (2007). *Analyzing qualitative data*. SAGE.

Goodwin, D., Pope, C., Mort, M., & Andrew, S. (2003). Ethics and ethnography: An experiential account. *Qualitative Health Research, 13*(4), 567–577. https://doi.org/10.1177/1049732302250723

Gotwals, A. W., & Cisterna, D. (2022). Formative assessment practice progressions for teacher preparation: A framework and illustrative case. *Teaching and Teacher Education, 110*, 1-13. https://doi.org/10.1016/j.tate.2021.103601

Gu, M. (2013). 《义务教育英语课程标准（2011 年版）》与《全日制义务教育普通高级中学英语课程标准（实验稿）》对比研究 [A comparative study of "English curriculum standards for compulsory education (2011)" and "English curriculum standards for full-time compulsory education and senior high schools (trial version)"]. *基础英语教育 [Journal of Basic English Education], 15*(1), 3-8.

Gu, Y. (2012). English curriculum and assessment for basic education in China. In J. Ruan & C. B. Leung (Eds.), *Perspectives on teaching and learning English literacy in China* (pp. 35-50). Springer.

Gu, 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

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105–117). SAGE.

Guba, E. G., & Lincoln, Y. S. (2005). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (3rd ed., pp. 191-215). SAGE.

Halliday, M. A. K. (1985). *An introduction to functional grammar*. Edward Arnold.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2000). *Construing experience through meaning : A language-based approach to cognition*. Continuum.

Hamlyn, D. (2005). Epistemology, History of. In T. Honderich (Ed.), *The Oxford companion to philosophy* (pp. 260-262). Oxford University Press.

Han, M., & Yang, X. (2001). Educational assessment in China: Lessons from history and future prospects. *Assessment in Education: Principles, Policy & Practice, 8*(1), 5-10. https://doi.org/10.1080/09695940120033216

Harding, L. (2014). Communicative language testing: Current issues and future research. *Language Assessment Quarterly, 11*(2), 186-197. https://doi.org/10.1080/15434303.2014.895829

Hargreaves, A. (2009). Preface. In K. D. Gariepy, B. L. Spencer, & J. C. Couture (Eds.), *Educational accountability: Professional voices from the field* (pp. xv–xviii). Sense Publishers.

Harlen, W. (2007). Criteria for evaluating systems for student assessment. *Studies in Educational Evaluation, 33*, 15-28. https://doi.org/10.1016/j.stueduc.2007.01.003

Harmer, J. (2007a). *How to teach English* (2nd ed.). Pearson Education.

Harmer, J. (2007b). *The practice of English language teaching* (4th ed.). Pearson Longman.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. https://doi.org/10.3102/003465430298487

Hayes, D. (2000). Cascade training and teachers' professional development. *ELT Journal, 54*(2), 135-145. https://doi.org/10.1093/elt/54.2.135

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

Howatt, A. P. R. (1987). From structures to communicative. *Annual Review of Applied Linguistics, 8*, 14-29. https://doi.org/10.1017/S0267190500000994

Huberman, A. M., & Miles, M. B. (1998). Data management and analysis methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Collecting and interpreting qualitative materials* (pp. 179-210). SAGE.

Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge University Press.

Hughes, J., & Sharrock, W. (1997). *The philosophy of social research* (3rd ed.). Longman.

Hui, S. K. F., Brown, G. T. L., & Chan, S. W. M. (2017). Assessment for learning and for accountability in classrooms: The experience of four Hong Kong primary school curriculum leaders. *Asia Pacific Education Review, 18*(1), 41–51. https://doi.org/10.1007/s12564-017-9469-6

Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: Selected readings* (pp. 269-293). Penguin.

Ingram, D. E. (1985). Assessing proficiency: An overview on some aspects of testing. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and assessing second language acquisition* (pp. 215-276). Multilingual matters.

Iphofen, R., & Tolich, M. (2018). Foundational issues in qualitative research ethics. In R. Iphofen & M. Tolich (Eds.), *The SAGE handbook of qualitative research ethics* (pp. 1-17). SAGE.

James, M. (2008). Assessment and learning. In S. Swaffield (Ed.), *Unlocking assessment: Understanding for reflection and application* (pp. 20-36). Routledge.

Jia, J. (2013). 2012 年深圳教育改革发展研究 [A study on education reform and development of Shenzhen in 2012]. In X. Zhang (Ed.), *深圳社会发展报告（2012-2013）[Annual report of social development of Shenzhen (2012-2013)]* (pp. 33-50). Social sciences academic press.

Jones, N., & Saville, N. (2016). *Learning oriented assessment: A systemic approach*. Cambridge University Press.

Karnieli-Miller, O., Strier, R., & Pessach, L. (2009). Power relations in qualitative research. *Qualitative Health Research, 19*(2), 279-289. https://doi.org/10.1177/1049732308329306

Katz, A., & Gottlieb, M. (2012). Assessment in the classroom. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (Vol. 1, pp. 161-169). Wiley-Blackwell.

Kennedy, A. (2005). Models of Continuing Professional Development: a framework for analysis. *Professional Development in Education, 40*(3), 336-351. https://doi.org/10.1080/19415257.2014.929293

Kerr, P. (2017). *Giving feedback on speaking. Part of the Cambridge Papers in ELT series.* Cambridge University Press. Retrieved from https://www.cambridge.org/elt/blog/wp-content/uploads/2018/02/Cambridge-Press_Whitepaper_Feedback_Speaking_2018.pdf

Koponen, M., & Riggenbach, H. (2000). Overview: Varying perspectives on fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 5–24). The University of Michigan Press.

Kowal, S., & O'Connell, D. C. (2014). Transcription as a crucial step of data analysis. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 64-78). SAGE.

Kumaravadivelu, B. (2006). TESOL methods: Changing tracks, challenging trends. *TESOL Quarterly, 40*(1), 59-81. https://doi.org/10.2307/40264511

Kvale, S. (1989). To validate is to question. In S. Kvale (Ed.), *Issues of validity in qualitative research* (pp. 73-92). Studentlitteratur.

Kvale, S. (1994). Ten standard objections to qualitative research interviews. *Journal of Phenomenological Psychology, 25*(2), 147–173. https://doi.org/10.1163/156916294X00016

Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing*. SAGE.

Kvale, S. (2007). *Doing interviews*. SAGE.

Lado, R. (1957). Sentence structure. *College Composition and Communication, 8*(1), 12-16. https://doi.org/10.2307/354410

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* (Vol. 2, pp. 756-771). SAGE.

Lantolf, J., & Thorne, S. (2006). Sociocultural theory and second language learning. In B. V. Patten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 201-224). Lawrence Erlbaum.

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

Leong, W. S., Ismail, H., Costa, J. S., & Tan, H. B. (2018). Assessment for learning research in East Asian countries. *Studies in Educational Evaluation, 59*, 270-277. https://doi.org/10.1016/j.stueduc.2018.09.005

Leung, C. (2004). Developing formative teacher assessment: Knowledge, practice, and change. *Language Assessment Quarterly, 1*(1), 19-41. https://doi.org/10.1207/s15434311laq0101_3

Leung, C. (2014). Classroom-based assessment issues for language teacher education. In A. J. Kunnan (Ed.), *The companion to language assessment* (Vol. 3, pp. 1510–1519). John Wiley & Sons, Inc.

Leung, C. (2020). Learning-oriented assessment: More than the chalkface. In M. E. Poehner & O. Inbar-Lourie (Eds.), *Toward a reconceptualization of second language classroom assessment* (pp. 85-106). Springer.

Leung, C. (2022). Language proficiency: From description to prescription and back? *Educational Linguistics, 1*(1), 56–81. https://doi.org/10.1515/eduling-2021-0006

Levi, T., & Inbar-Lourie, O. (2020). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly, 17*(2), 168-182. https://doi.org/10.1080/15434303.2019.1692347

Levy-Vered, A., & Alhija, F. N.-A. (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

Li, L. (2013). The complexity of language teachers' beliefs and practice: One EFL teacher's theories. *The Language Learning Journal, 41*(2), 175-190. https://doi.org/10.1080/09571736.2013.790132

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning, 60*(2), 309-365. https://doi.org/10.1111/j.1467-9922.2010.00561.x

Li, X. (1990). How powerful can a language test be? The MET in China. *Journal of Multilingual & Multicultural Development, 11*(5), 393-404.

Lightbown, P. M., & Spada, N. (2013). *How languages are learned* (4th ed.). Oxford University Press.

Lin, M. (2015). "Bad students go to vocational schools!": Education, social reproduction and migrant youth in urban China. *The China Journal, 73*(1), 108-131. https://doi.org/10.1086/679271

Littlewood, W. (2011). Communicative language teaching: An expanding concept for a changing world. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2, pp. 541-557). Routledge.

Liu, D. (2008). 英语课程从曲折走向有序 [The twists and turns of the English curriculum]. *基础教育外语教学研究 [Foreign Language Teaching and Research in Basic Education], 89*, 3-7.

Liu, F., & Maitlis, S. (2010). Nonparticipant observation. In A. J. Mills, G. Durepos, & E. Wiebe (Eds.), *Encyclopedia of case study research* (pp. 610-611). SAGE.

Liu, J., & Xu, Y. (2017). Assessment for learning in English language classrooms in China: Contexts, problems, and solutions. In H. Reinders, D. Nunan, & B. Zou (Eds.), *Innovation in language learning and teaching: The case of China* (pp. 17-37). Palgrave Macmillan.

Liu, W., & Wang, Q. (2020). Walking with bound feet: Teachers' lived experiences in China's English curriculum change. *Language, Culture and Curriculum, 33*(3), 242-257. https://doi.org/10.1080/07908318.2019.1615077

Liu, Y., Mishan, F., & Chambers, A. (2021). Investigating EFL teachers' perceptions of taskbased language teaching in higher education in China. *The Language Learning Journal, 49*(2), 131-146. https://doi.org/10.1080/09571736.2018.1465110

Lopes, J., & Oliveira, C. (2017). Classroom discipline: Theory and practice. In J. P. Bakken (Ed.), *Classrooms: Academic content and behavior strategy instruction for students with and without disabilities* (pp. 231-253). Nova Science Publishers.

Ma, M., & Bui, G. (2021). Chinese secondary school teachers' conceptions of L2 assessment: A mixed-methods study. *Studies in Second Language Learning and Teaching, 11*(3), 445-472. https://doi.org/10.14746/ssllt.2021.11.3.7

Mansell, W., James, M., & Assessment Reform Group. (2009). *Assessment in schools: Fit for purpose? A commentary by the teaching and learning research programme*. Economic and Social Research Council, Teaching and Learning Research Programme.

McDonough, J., Shaw, C., & Masuhara, H. (2013). *Materials and methods in ELT: A teacher's guide* (3rd ed.). Wiley-Blackwell, A John Wiley & Sons Ltd, Publication.

McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice, 22*(4), 34-43. https://doi.org/10.1111/j.1745-3992.2003.tb00142.x

McNamara, T. (1996). *Measuring second language performance*. Longman.

McNamara, T. (2013). Language testing: History, validity, policy. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology* (Vol. 1, pp. 341-352). American Psychological Association.

Meng, W. (2009). 激励教学法在初中英语教学中的运用 [The application of stimulation teaching method in junior English teaching]. *基础教育 [Journal of Schooling Studies], 6*(12), 58-60.

Min, Z., He, L., & Zhang, J. (2020). Review of recent empirical research (2011–2018) on language assessment in China. *Language Teaching, 53*, 316-340. https://doi.org/10.1017/S0261444820000051

Ministry of Education. (1999). *新中国教育五十年 [Education in fifty years of the People's Republic of China]*. People's Education Press.

Ministry of Education. (2001). *全日制义务教育普通高级中学英语课程标准（实验稿）[English curriculum standards for full-time compulsory education and senior high schools (trial version)]*. Beijing Normal University Press.

Ministry of Education. (2011). *义务教育英语课程标准 [English curriculum standards for compulsory education]*. Beijing Normal University Press.

Ministry of Education. (2022). *义务教育英语课程标准（2022 年版） [English curriculum standards for compulsory education (2022)]* Beijing Normal University Publishing Group.

Mitchell, R., & Myles, F. (2004). *Second language learning theories* (2nd ed.). Arnold; Distributed in the United States of America by Oxford University Press.

Molla, T., & Nolan, A. (2020). Teacher agency and professional practice. *Teachers and Teaching: Theory and Practice, 26*(1), 67-87. https://doi.org/10.1080/13540602.2020.1740196

Moon, K., & Blackman, D. (2014). A guide to understanding social science research for natural scientists. *Conservation Biology, 28*, 1167-1177. https://doi.org/10.1111/cobi.12326

Moses, J. W., & Knutsen, T. (2012). *Ways of knowing: Competing methodologies in social and political research* (2nd ed.). Hampshire, United Kingdom.

National Bureau of Statistics. (2020). *主要城市年度数据 [Annual statistics for major cities]*. Retrieved from https://data.stats.gov.cn/easyquery.htm?cn=E0105&zb=A01®=440300&sj=2020 (in Chinese)

Näykki, P., Kontturi, H., Seppänen, V., Impiö, N., & Järvelä, S. (2021). Teachers as learners – a qualitative exploration of pre-service and in-service teachers continuous learning community OpenDigi. *Journal of Education for Teaching, 47*(4), 495-512. https://doi.org/10.1080/02607476.2021.1904777

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199-218. https://doi.org/10.1080/03075070600572090

Nikander, P. (2008). Working with transcripts and translated data. *Qualitative Research in Psychology, 5*, 225–231. https://doi.org/10.1080/14780880802314346

Nusche, D. (2016). Student assessment and its relationship with curriculum, teaching and learning in the twenty-first century. In D. Wyse, L. Hayward, & J. Pandya (Eds.), *The SAGE handbook of curriculum, pedagogy and assessment* (Vol. 2, pp. 838-852). SAGE.

Ohta, A. (2000). Rethinking interaction in SLA: Developmentally appropriate assistance in the zone of proximal development and the acquisition of L2 grammar. In J. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 51-78). Oxford University Press.

Oller, J. W. (1979). *Language tests at school: A pragmatic approach*. Longman.

Opdenakker, R. (2006). Advantages and disadvantages of four interview techniques in qualitative research. *Forum: Qualitative Social Research, 7*(4), http://www.qualitative-research.net/index.php/fqs/article/view/175/391#g121.

Outhwaite, W. (1986). *Understanding social life: The method called verstehen* (2nd ed.). Jean Stroud.

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. John Benjamins Publishing Company.

Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting second language development*. Springer.

Poehner, M. E., & Lantolf, J. (2023). Advancing L2 dynamic assessment: Innovations in Chinese contexts. *Language Assessment Quarterly, 20*(1), 1-19. https://doi.org/10.1080/15434303.2022.2158465

Priestley, M., Biesta, G., & Robinson, S. (2015). *Teacher agency: An ecological approach*. Bloomsbury.

Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing, 22*(5), 142–173.

Qian, D. D. (2014). School-based English language assessment as a high-stakes examination component in Hong Kong: Insights of frontline assessors. *Assessment in Education: Principles, Policy & Practice, 21*(3), 251–270. https://doi.org/10.1080/0969594X.2014.915207

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

Richards, J. (2006). *Communicative language teaching today*. Cambridge University Press.

Rogoff, B. (1997). Evaluating development in the process of participation: Theory, methods, and practice building on each other. In E. Amsel & K. A. Renninger (Eds.), *Change*

and development: Issues of theory, method, and application (pp. 265-285). Lawrence Erlbaum Associates.

Rowe, W. (2014). Positionality. In D. Coghlan & M. Brydon-Miller (Eds.), *The SAGE encyclopedia of action research* (pp. 628). SAGE.

Rowlands, J. (2021). Interviewee transcript review as a tool to improve data quality and participant confidence in sensitive research. *International Journal of Qualitative Methods, 20*, 1-11. https://doi.org/10.1177/16094069211066170

Rubin, H., & Rubin, I. (2005). *Qualitative interviewing: The art of hearing data* (2nd ed.). SAGE.

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education, 35*(5), 535-550. https://doi.org/10.1080/02602930903541015

Sato, K., & Kleinsasser, R. (1999). Communicative language teaching (CLT): Practical understandings. *The Modern Language Journal, 83*(4), 494-517. https://doi.org/10.1111/0026-7902.00037

Savignon, S. J. (1972). Teaching for communicative competence: A research report. *Audio-Visual Language Journal, 10*(3), 153-162.

Savignon, S. J. (1991). Communicative language teaching: State of the art. *TESOL Quarterly, 25*(2), 261-277. https://doi.org/10.2307/3587463

Saville, N. (2021). Learning-oriented assessment: Basic concepts and frameworks in using assessment to support language learning. In A. Gebril (Ed.), *Learning-oriented language assessment: Putting theories into practice* (pp. 13-33). Routledge.

Schissel, J. L., Leung, C., & Chalhoub-Deville, M. (2019). The construct of multilingualism in language testing. *Language Assessment Quarterly, 16*(4-5), 373-378. https://doi.org/10.1080/15434303.2019.1680679

Schissel, J. L., Morales, J., Leung, C., & López-Gopar, M. (2019). Insights and reflections on multilingual approaches to assessment. *Bilingual Basics*. http://newsmanager.commpartners.com/tesolbeis/issues/2019-07-08/2.html

Schwandt, T. A. (2000). Three epistemological stances for qualitative inquiry: Interpretivism, hermeneutics, and social constructivism. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 189–214). SAGE.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39-83). Rand McNally.

Shenzhen Admission and Examination Office. (2020). *我市 2020 年中考成绩公布 [Announcement of Zhongkao score 2020 in Shenzhen]*. Retrieved from http://szeb.sz.gov.cn/szzkw/wx/zk/content/post_8990094.html

Shenzhen Admission and Examination Office, & Shenzhen Institute of Education Sciences. (2021). *深圳市初中毕业生学业考试说明 2021 [The guidelines for middle school graduation tests in Shenzhen 2021]*. Southern Publishing Media.

Shenzhen Education. (2020). *《深圳市初中学生综合素质基本指标评价标准》意见征集情况 [Report for comments on the evaluation criteria of Comprehensive Assessment for middle school students in Shenzhen]*. Retrieved from http://szeb.sz.gov.cn/home/gzcy/zjdc/jgfk/content/post_7570802.html

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4-14. https://doi.org/10.3102/0013189X029007004

Shi, J. (2001). *我国基础教育课程政策发展变化的历史轨迹 [Tracing the historical development of curriculum policies in China's basic education]*. Retrieved October 29, 2018 from http://www.edu.cn/edu/ji_chu/zong_he/ke_cheng/gai_ge/200603/t20060323_19001.shtml

Shi, L. (2006). The successors to Confucianism or a new generation? A questionnaire study on Chinese students' culture of learning English. *Language, Culture and Curriculum, 19*(1), 122-147. https://doi.org/10.1080/07908310608668758

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings *Language Learning, 49*(1), 93-120. https://doi.org/10.1111/1467-9922.00071

Skinner, B. F. (1957). *Verbal behavior*. Appleton.

Smith, L. M. (1978). An evolving logic of participant observation, educational ethnography, and other case studies. In L. Shuman (Ed.), *Review of Research in Education* (Vol. 6, pp. 316-377). Peacock.

Song, S. (2005). 建国以来我国高考制度的变革与发展 [Transform and development of college entrance examination system since 1949]. *清华大学教育研究 [Tsinghua Journal of Education], 26*, 103-108 (in Chinese).

Spolsky, B. (2008). Language assessment in historical and future perspectives. In E. Shohamy & N. H. Homberger (Eds.), *Encyclopedia of language and education* (2nd ed., Vol. 7, pp. 445-454). Springer.

Stake, R. E. (1994). Case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 236-247). SAGE.

Stake, R. E. (1995). *The art of case study research*. SAGE.

Steiner-Khamsi, G. (2014). Cross-national policy borrowing: Understanding reception and translation. *Asia Pacific Journal of Education, 34*(2), 153-167. https://doi.org/10.1080/02188791.2013.875649

Stobart, G. (2012). Validity in formative assessment. In J. Gardner (Ed.), *Assessment and learning* (2nd ed., pp. 133 – 146). SAGE.

Sun, M. (2016). 多元形成性评价在小学英语教学中的有效应用 [The effective application of formative assessment in primary school English teaching]. *中国教育学刊 [Journal of The Chinese Society of Education], 12*, 8-12.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Newbury House.

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

Tavakoli, P., & Hunter, A.-M. (2018). Is fluency being 'neglected' in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research, 22*(3), 330–349. https://doi.org/10.1177/1362168817708462

Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal, 21*(5), 615–631. https://doi.org/10.1080/01411920120095780

Turner, F., Brownhill, S., & Wilson, E. (2017). The transfer of content knowledge in a cascade model of professional development. *Teacher Development, 21*(2), 175–191. https://doi.org/10.1080/13664530.2016.1205508

Vickers, E., & Zeng, X. (2017). *Education and society in post-Mao China*. Routledge.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Walsh, S. (2011). *Exploring classroom discourse: Language in action*. Routledge.

Wang, H. (2008). 关于课堂教学评价问题的省思 [Reflection on classroom assessment]. *河北农业大学学报（农林教育版）[Journal of Agricultural University of Hebei (Agriculture and Forestry Education Edition)], 10*(2), 142-145.

Wang, Q. (1999). 面向 21 世纪义务教育阶段外语课程改革的思考与意见 [Some thoughts on the 21st century foreign language curriculum reform for nine-year compulsory education]. *基础教育外语教学研究 [Foreign Language Teaching & Research in Basic Education], 2*, 46-50.

Wang, Q. (2007). The national curriculum changes and their effects on English language teaching in the People's Republic of China. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching* (pp. 87-105). Springer.

Wang, Q., & Chen, Z. (2012). Twenty-first century senior high school English curriculum reform in China. In J. Ruan & C. B. Leung (Eds.), *Perspectives on teaching and learning English literacy in China* (pp. 85-103). Springer.

Wang, Z. (2016). 科举考试制度的全景敞视 [A panopticon on the imperial examination system]. *课程教学研究 [Journal of Curriculum and Instruction], 4*, 8-13.

Waring, M. (2017). Finding your theoretical position. In R. Coe, M. Waring, L. V. Hedges, & J. Arthur (Eds.), *Research methods & methodologies in education* (2nd ed., pp. 15-22). SAGE.

Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Erlbaum.

Williams, J. P. (2008). Nonparticipant observation. In L. M. Given (Ed.), *SAGE encyclopedia of qualitative research methods* (Vol. 2, pp. 561-562). SAGE.

Wilson, J., Ahrendt, C., Fudge, E. A., Raiche, A., Beard, G., & MacArthur, C. (2021). Elementary teachers' perceptions of automated feedback and automated scoring: Transforming the teaching and learning of writing using automated writing evaluation. *Computers & Education, 168*, 1-11. https://doi.org/10.1016/j.compedu.2021.104208

Xiao, Y., & Yang, M. (2019). Formative assessment and self-regulated learning: How formative assessment supports students' self-regulation in English language learning. *System, 81*, 39-49. https://doi.org/10.1016/j.system.2019.01.004

Xu, Y., & Harfitt, G. J. (2019). Is assessment for learning feasible in large classes? Challenges and coping strategies from three case studies. *Asia-Pacific Journal of Teacher Education, 47*(5), 472-486. https://doi.org/10.1080/1359866X.2018.1555790

Xu, Y., & He, L. (2019). How pre-service teachers' conceptions of assessment change over practicum: Implications for teacher assessment literacy. *Frontiers in Education, 4*, 1-16. https://doi.org/10.3389/feduc.2019.00145

Yan, C. (2012). 'We can only change in a small way': A study of secondary English teachers' implementation of curriculum reform in China. *Journal of Educational Change, 13*, 431–447. https://doi.org/10.1007/s10833-012-9186-1

Yan, C. (2015). 'We can't change much unless the exams change': Teachers' dilemmas in the curriculum reform in China. *Improving Schools, 18*(1), 5–19. https://doi.org/10.1177/1365480214553744

Yan, X., Zhang, C., & Fan, J. (2018). "Assessment knowledge is important, but …": How contextual and experiential factors mediate assessment practice and training needs of language teachers. *System, 74*, 158-168. https://doi.org/j.system.2018.03.003

Yin, R. (2012). *Applications of case study research* (3rd ed.). SAGE.

Yin, R. (2014). *Case study research: Design and methods* (5th ed.). SAGE.

Yin, R. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE.

Zhang, C., & Shu, Y. (Eds.). (2013). *英语：教师用书（八年级上册） [English: Teacher guidebook (grade 8 volume 1)]*. Shanghai Educational Publishing House.

Zhang, D. (2012). Chinese primary school English curriculum reform. In J. Ruan & C. B. Leung (Eds.), *Perspectives on teaching and learning English literacy in China* (pp. 67-83). Springer.

Zhang, F., & Liu, Y. (2014). A study of secondary school English teachers' beliefs in the context of curriculum reform in China. *Language Teaching Research, 18*(2), 187–204. https://doi.org/10.1177/1362168813505940

Zhang, J. (2013). 中国高中英语教育形成性评价实施情况研究 [A study of formative assessment in senior high school English teaching]. *课程·教材·教法 [Curriculum, Teaching Material and Method], 33*(9), 60-67.

Zhang, Y., & Xi, J. (2023). Fostering self-regulated young writers: Dynamic assessment of metacognitive competence in secondary school EFL class. *Language Assessment Quarterly, 20*(1), 88-107. https://doi.org/10.1080/15434303.2022.2103702

Zhao, D. (2013). 教学中的形成性评价：是什么及如何推进 [Formative assessment in teaching: What it is and how to improve it]. *教育科学研究 [Educational Science Research], 3*, 47-51.

Zheng, R., & Xu, D. (2020). 高考科目改革向何处去 ——基于 70 年高考科目设置变迁与困境的分析[Where is the reform of college entrance examination subjects heading for? Analysis of the Changes and Predicaments in College Entrance Examination Subject Setting in the Past 70 Years]. *复旦教育论坛[Fudan Education Forum], 18*(3), 91-97.

Zheng, X., & Borg, S. (2014). Task-based learning and teaching in China: Secondary school teachers' beliefs and practices. *Language Teaching Research, 18*(2), 205-221. https://doi.org/10.1177/1362168813505941

Appendix 1 Observation scheme (sample)

Observation scheme

School/teacher: A.T1 Class period: L2 Number of students: 48 Date/time: Sep 15 9:35

Unit 2 Reading passage

Time (e.g.: 10:45)	What (topic, vocabulary, grammar, etc.)	How (question & answer, teacher checking, exercise, group work, individual work etc.)	What assessment opportunities happened
9:34	Lead-in	T: what do you do in your free time (to all Ss) (ppt says free talk) S: homework, TV, video game, football, iPhone, read books T: Do you play chess S: No! T: Do you know how many squares are there on a chessboard? S: 64 T: How do you know? S: count (Chinese)	arise interest, prior knowledge? (very quick, 2mins)
9:35	Lead-in: picture	T: who are the people in the picture? Ss choose. T: What are they doing? Ss choose. T: Where did the story happen? Ss choose. T: how do you know? 1 S says "from their clothes", 1 S says "from the building", 1 S says "from the book"	Multiple choices from Reading: part B Offer opportunity for Ss to speak
9:37	Read the passage	T asks Ss to read the passage after the recording (one sentence at a time). T corrects Ss pronunciation. T asks Ss "can you understand the reading passage", Ss say Yes. T asks "are there any new words for you?" 1 Ss says "just something" T asks again, some student say Yes T ask Ss to read the passage after the recording again (one sentence at a time)	Correct pronunciation Check understand and learning gap
9:46	Answer the questions	T asks 1 S "what was the king's favourite game". S answers "Chess". T says "can you answer in a full sentence". S answers "the king's favourite game was chess". T asks 1 S "what prize did the old man want". S answers "Rice". T says "can you answer in a full sentence". S answers "he wanted the rice". T asks 1 S "how many grains of rice should the king put on the fifth square of the chess board". S answers "sixteen", but T doesn't ask for a full	Q&A, instant feedback, criteria emphasising; Q&A, instant feedback; Q&A

		sentence and let her sit down. T asks 1 S "what was the king's problem". S couldn't give an answers immediately. T asks "where can you find the answer", S finally answers "he did not have enough rice for all the squares." T reads the original text of the passage. T asks 1 S "what do you think of the old man in the passage", S couldn't answer, T says the answer "the old man was wise". Other Ss says "foolish" or "clever", but T didn't respond.	Q&A, scaffolding Q&A
9:50	True or False	T picks 1S, "the story happened in antient Italy, T/F", S says F. T asks "how to correct it". S didn't answer, other Ss answer. T says the answer "…in India" T picks 1S, "the old man challenge the king to a game, T/F". S didn't answer and didn't know how to correct the sentence. Other Ss answer, T shows the answer T picks 1S, True.	Q&A Q&A Q&A
9:53	Correct order	T asks Ss to re-read the passage and find the correct order for the sentences. Some Ss answer out loud.	Q&A
9:54	Complete the table (p21)	T asks Ss to read the story again, then complete the table on the textbook. T: if finish, put up your hand (few Ss put up, even though they've finished, perhaps afraid of being picked up?) (T gives feedback for some Ss while they do the exercise) T check the answers altogether (T didn't check each S, only the Ss sitting next to her). T reads the information in the table and let Ss tell their answers. T asks "why can't we just write 'old man', elicit Ss answers, review the grammar knowledge.	A way of gaining info from Ss? Instant feedback Group feedback S-T feedback
10:01	Question	T: How many grains of rice should be put on the last square? Ss answer in Chinese, T shows the answer in number. (Reading: D3)	Q&A, feedback,
10:02	Key phrases	T asks Ss to take notes on important phrases in the passage. T gives the Chinese meaning of the phrases, asks Ss to translate and highlight in the text. T asks for each phrase, only some Ss respond, then T shows the answers on ppt.	Q&A (not effective though)

10:06	Role play	(Ss sit at the back losing attention) T asks how to say 两粒米 “two grains of rice”. Ss answers. T reminds Ss of other meanings of specific words. T asks Ss to role-play the passage. Ss ask “加分吗” “will you add to our points” 1st S makes mistakes, other Ss correct her Another two groups of Ss	Q&A Prior knowledge, check Ss memory, review 加分 incentive Peer assessment /feedback

Appendix 2 Before-observation interview schedule

1. 基本情况

1. Background Information

a) 关于您：

a) About yourself:

- 请您简单介绍一下自己。
- Could you introduce yourself a little?
- 您从事教学工作多长时间了？
- How many years have you been teaching?
- 在您成为教师之前，您有过哪些英语教育相关的教育经历或培训经历？
- What was your academic or training background related to English education before becoming a teacher?

b) 关于您的学生和学校：

b) About your students and school:

- 可以简单介绍一下您的学生的英语水平和英语学习情况吗？
- How would you describe your students' level of English?
- 可以简单介绍一下您的学校的基本情况吗？
- Could you describe your school a little?

2. 课堂教学与评价

2. About learning, teaching and assessment in your classroom

a) 备课：

a) How do you plan for your class?

- 您会借助哪些教学资源备课呢？
- What teaching materials do you use to plan your lessons?
- 可以详细说说，您是如何使用这些资源备课的吗？
- Could you explain in detail how you use these materials to plan your lessons?

b) 评价：

b) Tell me about your assessment practices:

- 在近期的教学中，您最常使用的评价方式是什么？可以举几个例子吗？
- Which assessment practice do you perform the most in the last semester/unit? Could you give me some examples?
- 请您依次简单地评论以下几种课堂评价方式：书面练习，课堂问答，形成性评价，自评/学生互评。
- Can you briefly comment on the following classroom assessment practices: written exercises, question and answer, formative assessment, self/peer assessment?
- 您最喜欢哪种类型的评价方式？为什么？
- What kinds of classroom assessment do you prefer? Why?

c) 评价、课本与教师用书：

c) About assessment, textbook and teacher guidebook

- 您使用的课本和教师用书（或其他的教学资源）会在课堂评价方面提供一些建议或参考案例吗？能否举一些例子？
- Do textbook and teacher guidebook (or other teaching materials) provide any suggestions or examples about classroom assessment tasks or process? Could you give me some examples?
- 您会参考或使用这些建议和参考案例吗？能否举一些您在教学实践中的例子？
- If yes, do you refer to or use these suggestions and examples? Could you give me some examples?
- 您觉得这些建议和参考案例对您的帮助大吗？
- To what extent do you find these suggestions and examples useful?

d) 形成性评价：

d) About formative assessment in practice

- 您会在教学过程中采用形成性评价吗？如果会的话，可以分享一些例子吗？如果不会的话，为什么？
- Do you use formative assessment in class? If yes, could you describe an occasion when you used formative assessment in your classroom? If no, could you explain why?
- 如果您在课堂中使用形成性评价的话，这种评价方式会对学生的学习有什么样的影响呢？
- If you use formative assessment in class, how does it influence your students' learning?
- 如果您在课堂中使用形成性评价的话，这种评价方式会对您的教学有什么样的影响呢？
- If you use formative assessment in class, how does it influence your teaching?
- 您觉得在中国，大多数人会如何定义评价这个词？为什么？
- How do you believe most people in China would define assessment?
- 您觉得，相比起中国传统意义上的评价，形成性评价的概念有什么不同或相似之处？
- To what extent do you think formative assessment is different from the conventional beliefs on assessment in China? Why?

3. 专业学习经历

3. About Professional Learning

- 在您的教育经历或培训经历中，您是否有接触过课堂评价方面的知识或培训？可以描述一下这方面的经历吗？
- In your academic or training experience, have you had any professional learning opportunities in assessment? Can you describe your experience?
- 在之前的教育经历或培训经历中，您经历过哪些类型的课堂评价呢？
- In your previous academic or training experience, what kind of classroom assessment did you experience?
- 这些经历对您的教学和评价有哪些影响？
- Have these experiences influence your teaching and assessment?
- 您会参考其他老师的课堂评价活动吗？可以举一些例子吗？

- Would you refer to other teachers' classroom assessment activities? Could you give me some examples?
- 您觉得评价的主要目的是什么？
- What do you think is the main purpose of assessment?

4. 教育评价政策：

4. About assessment policy:

- 哪些教育评价政策对您的工作有影响？（可以从国家政策，地区政策和学校政策三个层面说）
- What assessment policy(ies) affect/s your work? (national, regional or school level)
- 可以分别举例描述一下这些政策对您的教学的影响吗？
- Can you give me an example of the way in which a particular policy (or several policies) has influenced your practice?
- 请您从英语学科的角度简单介绍一下目前在深圳推行的中考改革。对您来说，目前最大的变化是什么？
- I understand that recently, a new Zhongkao reform is being implemented in Shenzhen. Can you tell me anything about it? What do you think are the major changes?
- 这些变化会如何影响您和您的学生呢？
- How are these changes influencing you and your students?
- 您觉得《英语课程标准》（2011 版）对您的工作有什么影响和帮助吗？
- How helpful do you find the curriculum standards for English subject (2011)?
- 您会根据课程标准中给出的评价建议设计您的课堂评价吗？
- Do you plan your classroom assessment according to the curriculum standards?

Appendix 3 After-observation interview schedules

T1 课后访谈
T1 after-observation interview

1. 关于课程
1. About curriculum:
在这个单元（第二单元）的教学中，您希望达到的教学目的是什么？
What did you want the students to learn in unit 2?
2. 关于教学
2. About pedagogy:
您的学生对这个单元知识的掌握情况如何？
How well do you think your students have learned during this unit?
您是如何了解学生对这个单元知识的掌握情况？
How did you gather that information?
您觉得这个情况达到了您的预期吗？
Do you think this information matches your expectations?
3. 关于评价
3. About assessment:
我将和您分享几个您在课堂中的评价瞬间（1-9），请您就以下几个问题分享一下您的看法：
I will share nine observed moments of assessment in your class and hope you can share your thoughts about these assessment moments.

- 在这个评价瞬间，您的目的是什么？
- What was the purpose of this activity?
- 您为什么会使用或设计这种评价方式？
- Why did you choose or design such activity?
- 通过这种方式，您获取了哪些和学生学习相关的信息？您会如何使用这些信息？
- What did you find out about learning in your classroom from this activity? How did you use the information?

1. 在单词课上，老师在屏幕上展示例句，并点名学生起来翻译例句。
   In vocabulary lesson, you showed several sentences on the screen and asked students to translate the sentences.
2. 在阅读课的开头，老师向全体学生提问“what do you do in your free time?”
   At the beginning of reading lesson, you ask all students “what do you do in your free time?”
3. 在语法课上，老师点名几位同学起来回答问题：“how old are you?”，“how many people are there in your family?”
   In grammar lesson, you invited a few students and asked “how old are you?”, “how many people are there in your family?”
4. 在语法课上，老师提问“what’s the date today?”有些学生回答“Wednesday”，有些学生回“sixteen”。老师又用中文提问了一遍“今天几号”，学生回答“十六”，老师说“Yes, it is September 16th.”
   In grammar lesson, you asked “what’s the date today?” Some students answered “Wednesday”, other answered “sixteen”. You asked again in Chinese “what’s the date today”, and some students responded in Chinese “sixteen”. You said, “yes, it is September 16th.”

5. 在报纸阅读课上，老师邀请学生在黑板上写出自己的答案，并邀请其他的学生批改并写上自己的答案。
   In reading exercise lesson, you invited a few students to write their answers on the blackboard and asked other students to correct their answers.
6. 老师在课堂上讲解批改过的作业，并提问学生“有多少同学做对了/做错了”或“你们选/填的是什么”。
   You explained the homework exercises in class and asked students “how many of you get this right/wrong” or “what did you choose?”
7. 老师在课堂上反复强调应使用完整的句子回答问题。
   You emphasised in class several times that students should use complete sentences to answer questions.
8. 在 more practice 课上，老师在屏幕上分享了这节课的两个目标
   In “more practice” lesson, you shared the two objectives of the lesson.
9. 在单元结束时，老师要求学生做一个自我评估。
   At the end of the unit, you asked students to complete the self-assessment form.

…….……..… End of T1 interview ……...……

T2 课后访谈
T2 after-observation interview

1. 关于课程
1. About curriculum:
在这个单元（第二单元）的教学中，您希望达到的教学目的是什么？
What did you want the students to learn in unit 2?
2. 关于教学
2. About pedagogy:
您的学生对这个单元知识的掌握情况如何？
How well do you think your students have learned during this unit?
您是如何了解学生对这个单元知识的掌握情况？
How did you gather that information?
您觉得这个情况达到了您的预期吗？
Do you think this information matches your expectations?
3. 关于评价
3. About assessment:
我将和您分享几个您在课堂中的评价瞬间（1-9），请您就以下几个问题分享一下您的看法：
I will share nine observed moments of assessment in your class and hope you can share your thoughts about these assessment moments.
- 在这个评价瞬间，您的目的是什么？
- What was the purpose of this activity?
- 您为什么会使用或设计这种评价方式？
- Why did you choose or design such activity?
- 通过这种方式，您获取了哪些和学生学习相关的信息？您会如何使用这些信息？

- What did you find out about learning in your classroom from this activity? How did you use the information?

1. 在阅读课上，老师向全体学生提问“what does palace mean”，学生们集体回答。
   In reading lesson, you asked all students “what does palace mean”, and students answered together.
2. 在阅读课上，老师请学生用英文讲与数字有关的中国故事。
   In reading lesson, you aksed students to tell story about numbers in Chinese history in English.
3. 老师在课堂上反复强调应使用完整的句子回答问题。
   You emphasised in class several times that students should use complete sentences to answer questions.
4. 老师说一个中文句子，要求学生们用英文翻译。如果学生们觉得翻译起来有困难，老师会提供单词和句型方面的提示。
   You said Chinese sentences and asked students to translate them into English. If they had difficulty in doing that, you provided hints related to vocabulary and sentence structure.
5. 在讲解听说考试信息转述题时，老师向全体学生提问“who is in the story”。一个学生回答“他的朋友”（正确答案是 an old man），老师听到了这个错误答案后向全体学生提问“是他的朋友吗”，部分学生回答“不是，是一个老人”。
   While explain the “retell a story” task in mock listening-and-speaking test, you asked all students “who is in the story?” One student said “his friend” (correct answer is an old man). You then asked other students “is it his friend?” Some students said, “no, an old man”.
6. 老师点名一个学生起来回答问题。学生回答“it take place in India”。老师说“it …”，似乎在暗示学生的答案有误，学生很快反应过来并回答“it took place in India”。
   You asked a student to complete the reading comprehension task. The student said “it take place in India”. You said, “it …”, which seemed like you were indicating there was something wrong with the answer. Then the student realised and said, “it took place in India.”
7. 老师在上课前安排默写/听写任务
   You instructed students to do dictation tasks at the beginning of lessons.
8. 老师在课堂上讲解批改过的作业，并提问学生“有多少同学做对了/做错了”或“你们选/填的是什么”。
   You explained homework exercises in class and asked students “how many of you get this right/wrong” or “what did you choose?”
9. 在语法课上，老师邀请学生在黑板上写出数字的英文。老师在黑板上批改学生的作答，其他学生在台下作出反馈（正确/错误/应该怎么写）
   In grammar lesson, you invited students to write numbers in English on the blackboard. You then corrected their answers, while other students responded to your correction (right, wrong, or how the number should be spelled).

…….…….…. End of T2 interview ……...……

T3 课后访谈
T3 after-observation interview

1. 关于课程

1. About curriculum:

在这个单元（第四单元）的教学中，您希望达到的教学目的是什么？

What did you want the students to learn in unit 4?

2. 关于教学

2. About pedagogy:

您的学生对这个单元知识的掌握情况如何？

How well do you think your students have learned during this unit?

您是如何了解学生对这个单元知识的掌握情况？

How did you gather that information?

您觉得这个情况达到了您的预期吗？

Do you think this information matches your expectations?

3. 关于评价

3. About assessment:

我将和您分享几个您在课堂中的评价瞬间（1-8），请您就以下几个问题分享一下您的看法：

I will share nine observed moments of assessment in your class and hope you can share your thoughts about these assessment moments.

- 在这个评价瞬间，您的目的是什么？
- What was the purpose of this activity?
- 您为什么会使用或设计这种评价方式？
- Why did you choose or design such activity?
- 通过这种方式，您获取了哪些和学生学习相关的信息？您会如何使用这些信息？
- What did you find out about learning in your classroom from this activity? How did you use the information?

1. 在阅读课上，老师问：How do these inventions help us in daily life? 学生回答：Telephone can help us keep in touch with others anytime and anywhere. 老师问：Are you sure? Can you take a telephone anywhere with you? 学生改为回答：mobile phone.
   In reading lessons, you asked, “How do these inventions help us in daily life?” One student answered, “Telephone can help us keep in touch with others anytime and anywhere.” You asked, “Are you sure? Can you take a telephone anywhere with you?” The student then said, “mobile phone.”
2. 老师要求学生以小组为单位，互相问答。老师点名一对学生起来分享他们的对话
   You asked students to form groups and have conversations based on the given questions. You invited a pair of students to demenstrate their conversation.
3. 老师在课上安排听写、做练习任务，并在班里巡视
   You instructed students to do dictation and exercises in class and walk around the classroom while they did the tasks.
4. 老师在听写、练习任务后要求学生自我批改
   You asked students to self-mark their answers after dictations and exercises.
5. 老师用英文描述单词，要求学生在台下口头作答
   You described words in English and asked students to say the words.
6. 老师要求学生用完整句子作答
   You asked students to answer your questions in complete sentences.
7. 在写作课上，老师要求学生相互批改作文，并为学生提供 evaluation form
   In writing lesson, you asked students to evaluate their peers’ writing and provided an evaluation form for them.
8. 在写作课和 more practice 课上，老师展示范文和其他班学生的优秀作业

In writing lesson and “more practice” lesson, you demonstrated sample writing and writing from students of another class.

…….…….…. End of T3 interview ……...……

T4 课后访谈
T4 after-observation interview

1. 关于课程
1. About curriculum:
在这个单元（第四单元）的教学中，您希望达到的教学目的是什么？
What did you want the students to learn in unit 4?
2. 关于教学
2. About pedagogy:
您的学生对这个单元知识的掌握情况如何？
How well do you think your students have learned during this unit?
您是如何了解学生对这个单元知识的掌握情况？
How did you gather that information?
您觉得这个情况达到了您的预期吗？
Do you think this information matches your expectations?
3. 关于评价
3. About assessment:
我将和您分享几个您在课堂中的评价瞬间（1-10），请您就以下几个问题分享一下您的看法：
I will share ten observed moments of assessment in your class and hope you can share your thoughts about these assessment moments.

- 在这个评价瞬间，您的目的是什么？
- What was the purpose of this activity?
- 您为什么会使用或设计这种评价方式？
- Why did you choose or design such activity?
- 通过这种方式，您获取了哪些和学生学习相关的信息？您会如何使用这些信息？
- What did you find out about learning in your classroom from this activity? How did you use the information?

1. 在阅读课上，老师要求学生在笔记上自己造句
   In reading lesson, you asked students to make sentences about the words they learnt on their notebook.
2. 在阅读课上，老师提问“how to say ‘三千’？”学生集体回答 three thousand。老师提问“需要加一个 s 吗？”学生集体回答“不需要”
   In reading lesson, you asked, “how to say ‘three thousand’ in English”. Students answered together, “three thousand”. You asked, “should we add a ‘-s’?” Students answered together, “no!”
3. 老师要求学生用完整句子作答
   You asked students to answer your questions in complete sentences.
4. 老师用中文说句子，要求学生用刚学过的词汇翻译句子

You said sentences in Chinese and asked students to translate the sentences with the words they just learnt.

5. 老师点名学生起来回答 “Bell ____ (invent) the telephone”一题，学生回答 “invent”，老师说“invent? He did it before.” 其他学生提示正确答案，学生立刻改正回答 “invented”
   You asked one students to complete the sentence “Bell ____ (invent) the telephone”. The student said ‘invent”. You said, “invent? He did it before.” Other students provided the correct answer, and the students immediately changed his answer to “invented”.
6. 老师在课上安排听写、做练习任务，并在班里巡视
   You instructed students to do dictation and exercises in class and walk around the classroom while they did the tasks.
7. 老师在听写、练习任务后要求学生自我批改
   You asked students to self-mark their answers after dictations and exercises.
8. 在讲解练习时，老师分享容易选/做错的答案，并解释为什么这些答案是错误的
   In exercise lesson, you shared common mistakes made by students and explain why these answers were wrong.
9. 在写作课中，老师引导学生通过翻译已给出的句子写作文
   In writing lesson, you instructed students to translate given sentences to write an article.
10. 在写作讲解中，老师分享参考答案
    In writing lesson, you shared sample writing.

…….……..… End of T4 interview ……...……

Appendix 4 Analysis form (sample)

Analysis of observed assessment practices

School/teacher: A.T2 Class period: L7 Number of students: 48 Date/time: Sep 20 7:45 Unit 2 Dictation, "more practice", exercise explanation

Assessment activity	Participant organisation	Content (what the assessment is about)	Feedback (praise, correction, comment, written/oral feedback, etc.)	Code
T asks Ss "what's the syntactic function of 'difference'", Ss answer. T says "we need to use an adjective here, the answer is different".	T-Ss	Check understanding of grammar	T: confirm the answer, explain the grammar	TO-G
T asks Ss to finish a cloze for vocabulary in workbook. After finishing, T reads the cloze passage and pause before the blank, Ss answer, T explains.	T-Ss	Exercise; check the answer	T: confirm the answer, explain the answer	TW-V
T asks "what degree this word is", Ss answer "comparative", T asks "what is the superlative form of this word", Ss answer "add '-est'". T explains the "-er" and "-est" ending.	T-Ss	Exercise; check the answer	T: confirm the answer, explain the answer	TO-G
T asks Ss to finish a reading exercise in workbook. After finishing, T asks what the theme of the passage is, Ss answer.	T-Ss	Reading task; check understanding of the passage	T: confirm the answer	TW-RE

Appendix 5 Ethics approval

University of Glasgow

College of Social Sciences

9 January 2020

Dear Jingwei Song

Project Title: Assessing for learning in English language classrooms in Shenzhen, China

Application No: 400190068

The College Research Ethics Committee has reviewed your application and has agreed that there is no objection on ethical grounds to the proposed study. It is happy therefore to approve the project, subject to the following conditions:

- Project end date: _31/08/2022
- The data should be held securely for a period of ten years after the completion of the research project, or for longer if specified by the research funder or sponsor, in accordance with the University's Code of Good Practice in Research: (http://www.gla.ac.uk/media/media_227599_en.pdf)
- The research should be carried out only on the sites, and/or with the groups and using the methods defined in the application.
- Any proposed changes in the protocol should be submitted for reassessment as an amendment to the original application. The *Request for Amendments to an Approved Application* form should be used: http://www.gla.ac.uk/colleges/socialsciences/students/ethics/forms/staffandpostgraduateresearchstudents/

Yours sincerely,

Dr Muir Houston
College Ethics Officer

Muir Houston, Senior Lecturer
College of Social Sciences Ethics Officer
Social Justice, Place and Lifelong Education Research
University of Glasgow
School of Education, St Andrew's Building, 11 Eldon Street
Glasgow G3 6NH
0044+141-330-4699
Muir.Houston@glasgow.ac.uk

Appendix 6 Participant information sheet

University of Glasgow

College of Social Sciences

Participant Information Sheet

Title of Project: **Assessing for learning in English language classrooms in Shenzhen, China**

Researcher: **Jingwei Song 宋京蔚**

Supervisors: **Prof. Louise Hayward, Dr Sally Zacharias, University of Glasgow**

Hi! My name is Jingwei Song. I am a PhD student in the University of Glasgow, and I am inviting you to take part in this research project.

Before you decide it is important for you to understand why the research is being done and what it will involve. Please take your time to read the following information carefully and discuss it with others if you wish. Ask me if there is anything that is not clear or if you would like more information. Take your time to decide whether or not you wish to take part.

Thank you for reading this.

An introduction to the study:

- The purpose of this study is to understand how teachers assess students' learning processes in English language classrooms, and how teachers think of assessment and current assessment policy.
- The study will involve classroom observation, which with your permission will last around two weeks, and two interviews, one before the observation and one after. The two interviews will last 60 minutes and 30 minutes respectively. Both observations and interviews will be audio-recorded.
- Your participation should be entirely voluntary. You are entitled to withdraw at any time without prejudice and without any reason.
- The study could benefit you from the following perspectives: firstly, you will have the chance to reflect on assessment practices and assessment policy changes; secondly, you may use this experience to reflect on your future teaching and assessment practices.
- If you feel anxious or uncomfortable during the research, please note that I am not an inspector or supervisor. I am a learner, like any of your students, and wish to learn from you about your professional knowledge and experience.
- The data collected from you will only be used for my PhD thesis, publication in articles and conference papers. During the research, any electronic data will be stored on my computer with password protection and encryption. Any paper documents will be kept in a locked facility.
- To protect confidentiality, you will be referred to as a pseudonym in any publication arising from the research, and any personal data of yours will be destroyed at the end of the project. If you withdraw from the study, any data related to you will not be used in the project.

- Please note that it may be impossible to completely guarantee confidentiality due to my presence in class and colleagues potentially becoming aware as to who is participating.

- Please note that confidentiality will be maintained as far as it is possible, unless during our conversation I hear anything which makes me worried that someone might be in danger of harm, I might have to inform relevant agencies of this.
- This study is co-sponsored by the China Scholarship Council and the University of Glasgow.

This project has been considered and approved by the College Research Ethics Committee. If you have any further inquiry, please contact me via email: **j.song.1@research.gla.ac.uk**, or my supervisor, Professor Louise Hayward, email: **Louise.Hayward@glasgow.ac.uk**

For further information or to pursue acomplaint, you can contact the College of Social Sciences Ethics Officer, Dr Muir Houston, email: Muir.Houston@glasgow.ac.uk

_____End of Participant Information Sheet_____

University of Glasgow
College of Social Sciences

课题参与者告知书

课题名称：以“课堂评价”为视角的深圳初中英语课堂教学探索

研究人员：宋京蔚

指导老师：Professor Louise Hayward & Dr Sally Zacharias

归属单位：School of Education, University of Glasgow
格拉斯哥大学教育学院

尊敬的老师：

您好！我是英国格拉斯哥大学教育学院的在读博士生，目前从事我国中学阶段英语课堂教学与评价的相关研究。在得到贵校的批准后，我将于近期在贵校开展有关课题的实地调研活动。

在此，我诚挚地邀请您参与到我的研究中。这对我的课题有着直接作用与重要意义，您的工作经历、课堂实践、教学理念将予以我直观和丰富的研究数据，以帮助我更准确和深入地了解深圳初中英语课堂教学与评价的相关情况。

按照研究的伦理指导原则、中英两地的相关法律法规，您决定参加这项研究之前，需要被告知如下重要事宜。其涵盖了本课题的目的、流程、信息安全以及其他相关重要信息，请您仔细阅读。若有任何关于本课题的疑问或顾虑，我将予以详尽的解答。感谢您的理解与合作！

- **研究目的：**本项目旨在了解初中英语老师的课堂评价活动，以及老师们对课堂评价及相关教育评价政策的理解。

- **访谈流程：**本研究将涉及两种研究方法：课堂观察和教师访谈。课堂观察时间为一个单元的教学时间，具体课时将根据老师的课时安排与上课进度作相应调整。同时，研究需要进行两场访谈，一场在课堂观

- 察前，一场在课堂观察后，每场访谈的时间分别为 60 分钟和 30 分钟。听课和访谈的过程将被录音，方便后续研究工作。

- **自愿原则：**您的参与需是完全自愿的。您可以选择在研究过程中的任何时间无条件退出。此举不会影响到您与研究人员之间的关系，也不会影响您的任何权益。

- **友善交流：**我将以学生的身份和学习的态度来到您的课堂，向您学习专业知识和实践经验，在访谈中向您请教和交流，就像您的其他学生一样。如果在研究过程中感到任何不适，也请及时与我沟通。

- **减少干扰：**因需要对您的教学进行观察，我将会进入教室旁听。我将尽可能地采取必要手段减轻一切对正常课堂教学的干扰。如果我的观察对您的教学造成了困扰，请您与我及时沟通以便调整我旁听的策略，以便进一步减少可能的影响。

- **隐私保护：**为了保护您的隐私，您的姓名、年龄、职位等隐私个人信息将会在论文和出版物中被隐去，并在课题完成后彻底销毁。如果您在研究过程中选择退出，与您相关的任何数据都不会被收录，且已收录信息也将被确保删除。

- **数据安全：**您提供的研究数据将被用于我的博士论文、期刊论文和会议论文等学术出版物中。研究数据中所有涉及个人信息的内容都将被模糊化处理或被剥离。我将采用电子加密和物理加密等双重手段以确保信息储存的安全。只有在获得授权的情况下，导师和论文评审等研究人员才可进行有条件地查阅。

- **伦理审查：**本课题业已通过格拉斯哥大学伦理审查委员会的两轮审查，以排除可能的伦理问题和风险，并确保本研究符合中国、英国以及欧盟地区的相关法律与法规。

- **法律义务：**如果我在实地调研的过程中发现了涉及人身安全和危害公共安全的行为，我有义务按照有关法律法规报告给相关的主理机构。此项与保障个人隐私与信息安全不相冲突。

- **资助机构：**本研究课题由中国国家留学基金委与英国格拉斯哥大学联合资助，且上述机构不会从任何渠道获取您的个人信息。

如果您有任何关于研究课题的问题，欢迎通过微信（）或邮箱（j.song.1@research.gla.ac.uk）与我取得联系。如有其他问题可与我的导师Professor Louise Hayward （Louise.Hayward@glasgow.ac.uk）或格拉斯哥大学社会科学院伦理办公室主任Dr Muir Houston（Muir.Houston@glasgow.ac.uk）联络沟通。

Appendix 7 Teacher Consent Form

University of Glasgow
College of Social Sciences

Teacher Consent Form

Title of Project: **Assessing for learning in English language classrooms in Shenzhen, China**

Researcher: **Jingwei Song 宋京蔚**

Supervisors: **Prof. Louise Hayward, Dr Sally Zacharias, School of Education, University of Glasgow**

I confirm that I have read and understood the Participant Information Sheet for the above study and have had the opportunity to ask questions.

I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.

About research methods:

- I consent to interviews/observations being audio-recorded.
- I acknowledge that copies of transcripts will be returned to participants for verification.

About confidentiality:

- I acknowledge that participants will be referred to by pseudonyms in any publications arising from the research.

About data usage and storage:

- All names and other material likely to identify individuals will be anonymised.
- The material will be treated as confidential and kept in secure storage at all times.
- The material containing personal data will be destroyed once the project is complete.
- The material may be used in future publications, both print and online.
- I agree to waive my copyright to any data collected as part of this project.

I **agree** to take part in this research study ☐

I **do not agree** to take part in this research study ☐

Name of Participant: …… Signature: ……

Date: ……

Name of Researcher: …… Signature: ……

Date: ……

…… **End of consent form** ……

University of Glasgow
College of Social Sciences

受访同意书

课题名称：以“课堂评价”为视角的深圳初中英语课堂教学探索

研究人员：宋京蔚

指导老师：Professor Louise Hayward & Dr Sally Zacharias

归属单位：School of Education, University of Glasgow 格拉斯哥大学教育学院

我已经阅读了有关本项研究的介绍，并且有机会就此项研究与研究人员讨论并提出问题。我提出的所有问题都已得到满意的答复。

我自愿参与本研究，且知晓我可以在研究过程中的任何时间退出本研究，无需提供任何理由。我确认已阅读如下重要事宜，并有充足时间对此进行考虑。

关于研究方法：

- 我知晓并同意访谈和上课过程被录音。
- 我有权要求查看访谈的录音记录，并对记录的内容进行确认和修改。

关于隐私、信息使用与存储：

- 我的姓名、年龄、职位等可辨识的身份信息将会在相关论文和出版物中被隐去。
- 我的个人信息和提供的研究数据将被采用电子加密和物理加密的双重手段妥善保存，仅供研究人员、指导老师和论文评审在获得授权的情况下查阅。
- 我的个人信息将会在研究项目结束后被销毁。
- 我提供的研究数据将在未来被用于论文发表。
- 我同意放弃对在该项目中提供的研究数据的版权。

我同意参与本研究 ☐

我不同意参与本研究 ☐

受访者：……

日期：……

研究人员：……

日期：……

Appendix 8 Principal Consent Form

University of Glasgow
College of Social Sciences

Principal Consent Form

Title of Project: **Assessing for learning in English language classrooms in Shenzhen, China.**

Researcher: **Jingwei Song 宋京蔚**

Supervisors: **Prof. Catherine Doherty, Prof. Louise Hayward, School of Education, University of Glasgow.**

I confirm that I have read and understood the Participant Information Sheet for the above study and have had the opportunity to ask questions.

I understand that my school's participation is voluntary and that I am free to withdraw at any time, without giving any reason.

About participation:

- I consent that the researcher can conduct the study in my school.
- I acknowledge that classroom observations and interviews will be conducted with the teachers.

About confidentiality:

- I acknowledge that participants will be referred to by pseudonym in any publications arising from the research.

About data usage and storage:

- All names and other material likely to identify individuals will be anonymised.
- The material will be treated as confidential and kept in secure storage at all times.
- The material containing personal data will be destroyed once the project is complete.
- The material may be used in future publications, both print and online.
- I agree to waive my copyright to any data collected as part of this project.

I **agree** to my school taking part in this research study ☐

I **do not agree** to take part in this research study ☐

Name of Participant …… Signature ……

Date ……

Name of Researcher …… Signature ……

Date ……

University of Glasgow
College of Social Sciences

受访同意书

课题名称：以“课堂评价”为视角的深圳初中英语课堂教学探索

研究人员：宋京蔚

指导老师：Professor Louise Hayward & Dr Sally Zacharias

归属单位：School of Education, University of Glasgow 格拉斯哥大学教育学院

我已经阅读有关本项研究的介绍，并且有机会就此项研究与研究人员讨论并提出问题。我提出的所有问题都已得到满意的答复。

我校自愿参与本次调研，且在研究过程中保留退出的权利，无需提供任何理由。我确认已阅读如下重要事宜，并有充足时间对此进行考虑。

关于参与研究：

- 我同意该研究人员在本校进行调研。
- 我知晓该研究人员将在学校进行课堂观察并对老师进行访谈。

关于隐私、信息使用与存储：

- 我知晓老师姓名、年龄、职位等可辨识的身份信息将会在相关论文和出版物中被隐去。
- 相关个人信息和提供的研究数据将被采用电子加密和物理加密的双重手段妥善保存，仅供研究人员、指导老师和论文评审在获得授权的情况下查阅。
- 相关个人信息将会在研究项目结束后被销毁。
- 本校提供的研究数据将在未来被用于论文发表。
- 本校同意放弃对在该项目中提供的研究数据的版权。

同意参与本研究 ☐

不同意参与本研究 ☐

学校领导：……

日期：……

研究人员：……

日期：……

Appendix 9 Parent and Student Participant Information Sheet

Parent and Student Consent Form

Title of Project: **Assessing for learning in English language classrooms in Shenzhen, China**

Researcher: **Jingwei Song 宋京蔚**

Supervisors: **Prof. Catherine Doherty, Prof. Louise Hayward, School of Education, University of Glasgow**

I confirm that I have read and understood the Participant Information Sheet for the above study and have had the opportunity to ask questions.

I understand that my child's participation is voluntary and that we are free to withdraw our consent at any time, without giving any reason.

About research methods:

- I consent to classes being audio-recorded.

About confidentiality:

- I acknowledge that participants will be referred to by pseudonyms in any publications arising from the research.

About data usage and storage:

- All names and other material likely to identify individuals will be anonymised.
- The material will be treated as confidential and kept in secure storage at all times.
- Any personal data will be destroyed once the project is complete.
- The material may be used in future publications, both print and online.
- I agree to waive my copyright to any data collected as part of this project.

I **agree** for my child to take part in this research study ☐

I **do not agree** for my child to take part in this research study ☐

Name of Parent/carer: ……

Signature: …… Date: ……

I **agree** to take part in this research study ☐

I **do not agree** to take part in this research study ☐

Name of Student: …… Signature: ……

University of Glasgow
College of Social Sciences

受访同意书

课题名称：以“课堂评价”为视角的深圳初中英语课堂教学探索

研究人员：宋京蔚

指导老师：Professor Louise Hayward & Dr Sally Zacharias

归属单位：School of Education, University of Glasgow 格拉斯哥大学教育学院

我已经阅读有关本项研究的介绍，并且有机会就此项研究与研究人员讨论并提出问题。我提出的所有问题都已得到满意的答复。

我/我的孩子自愿参与本次调研，且在研究过程中保留退出的权利，无需提供任何理由。我确认已阅读如下重要事宜，并有充足时间对此进行考虑。

关于研究方法：

- 我同意课堂过程被录音。

关于研究方法：

- 我知晓研究人员参与者将在所有和该研究相关的论文和出版物中被隐去姓名。

关于隐私、信息使用与存储：

- 我知晓学生的姓名、年龄等可辨识的身份信息将会在相关论文和出版物中被隐去。
- 相关个人信息和提供的研究数据将被采用电子加密和物理加密的双重手段妥善保存，仅供研究人员、指导老师和论文评审在获得授权的情况下查阅。
- 相关个人信息将会在研究项目结束后被销毁。
- 我/我的孩子提供的研究数据将在未来被用于论文发表。
- 我/我的孩子同意放弃对在该项目中提供的研究数据的版权。

同意参与本研究 ☐

不同意参与本研究 ☐

受访者：……

日期：……

研究人员：……

日期：……

Appendix 11 Mock listening-and-speaking test

2021 年深圳市中考英语听说考试考前冲刺 2

（本测试共三大题，满分 25 分）

一、模仿朗读（共 4 分）

听以下短文一遍，你有 50 秒钟的准备时间，然后模仿朗读。当听到“开始录音”的信号后，请在 50 秒钟内朗读短文，当听到“停止录音”的信号时，停止朗读。

The first chocolate was eaten by people in South America hundreds of years ago. In those days, people did not really eat chocolate. They used cocoa bean to make a chocolate drink and they enjoyed it very much. Many years later, the cocoa bean was brought to other countries and people came to love the taste of chocolate.

二、信息获取（共 10 分）

第一节 听选信息（1×6=6 分）

听三段对话，每段播放两遍。各段播放前你有 10 秒钟的阅题时间。各段播放后有两个问题。请根据所听到的问题和括号内的提示，选择正确的信息，在 8 秒钟内口头回答。

听第一段对话，回答第 1-2 两个问题。现在你有 10 秒钟的阅题时间。

1. How does Jack usually go to school?
(On foot. / By bike. / By bus.)
2. What is the relationship between the two speakers?
(Mother and son. / Classmates. / Teacher and student.)

听第二段对话，回答第 3-4 两个问题。现在你有 10 秒钟的阅题时间。

3. Where does Judy come from?
(China. / Japan. / Korea.)
4. What does Mike think of making noise when eating noodles?
(It’s an impolite manner. / It means the noodles are bad. / It can show respect to the cook.)

听第三段对话，回答第 5-6 两个问题。现在你有 10 秒钟的阅题时间。

5. How old is Betty now?
(13. / 14. / 15.)
6. Where are the speakers talking?
(In a pet shop. / In Betty’s home. / In a park.)

第二节 回答问题（1×4=4 分）

听下面一段短文，播放两遍。请根据所听内容回答第 7-10 四个问题。当听到“开始录音”的信号后，请在 8 秒钟内口头回答。

7. When did Nick move to Shenzhen?
8. What do Nick's parents do?
9. What does the speaker think of Nick?
10. What does the speaker hope to do?

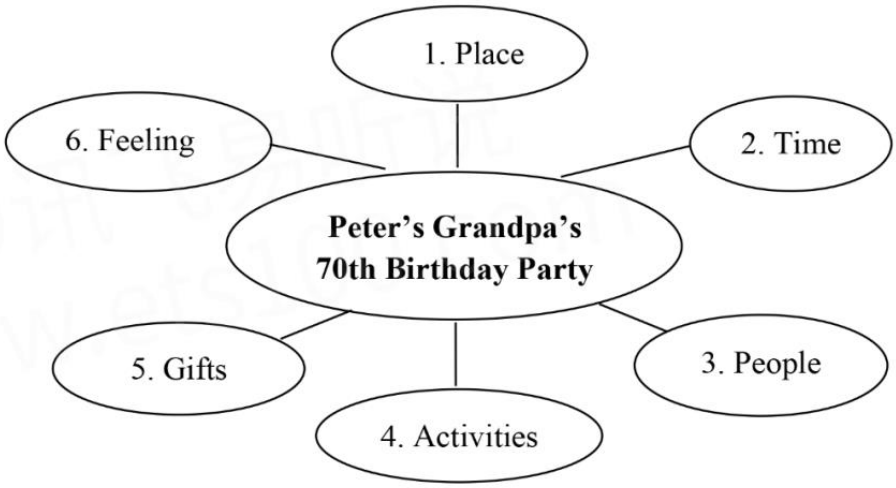
三、短文复述及提问（共 11 分）

第一节 短文复述（8 分）

你将听到 Peter 讲述他爷爷 70 岁寿宴的情况，录音播放两遍。请根据所听到的内容，依据思维导图的提示，在 60 秒钟内复述短文。

现在，你有 15 秒钟的时间阅读思维导图。你的复述可以这样开始：

It was Peter's grandpa's 70th birthday yesterday. …



下面你有 60 秒钟的时间准备。你的复述可以这样开始：

It was Peter's grandpa's 70th birthday yesterday. …

下面请听录音两遍。

第二节 提问（1.5×2=3 分）

你希望了解更多相关信息，请根据以下提示向 Peter 提两个问题。每个问题有 15 秒钟的准备时间和 8 秒钟的提问时间。

请你准备第一个问题。

11. 你想了解 Peter 的画是关于什么的，你问 Peter：

请你准备第二个问题。

12. 你想知道 Peter 是否和他祖父母住在一起，你问 Peter：

参考答案（范文）和录音原文：

一、模仿朗读（略）

二、信息获取

第一节 听选信息

参考答案：

1. Jack usually goes to school by bus.
2. They’re teacher and student.

录音原文：

W: You are late again, Jack! Why?
M: I’m sorry, Mrs. Smith. My bike broke down halfway, so I walked to school.
W: I remember you usually come to school by bus.
M: Yes. But I got up late today and missed the bus.

参考答案：

3. Judy comes from China.
4. It can show respect to the cook.

录音原文：

W: Hey, Mike. You had better not make so much noise when eating noodles.
M: Why?
W: Well, in our Chinese culture, we consider it an impolite manner.
M: I see, Judy. But in my home country Japan, it means you like the noodles. It’s also a way to show your respect to the cook.
W: Well, I see. Then that’s the same in Korea.

参考答案：

5. She is 15.
6. They’re talking in a park.

录音原文：

M: How long have you had your dog, Betty?
W: For two years. It was a gift for my 13th birthday.
M: Lucky you! My parents don’t allow me to have one.
W: Well, it’s not a bad thing. Pets often need much care.
M: You’re right. Oh, the park will close soon.
W: Let’s leave now.

第二节 回答问题

参考答案：

7. He moved there last summer.
8. They’re workers.

9. He’s smart and helpful.
10. She hopes to study in the same high school with Nick.

录音原文：

My friend Nick was born in Hunan. Last summer, he moved to Shenzhen with his parents and began to study at our school. His parents are workers in a factory and always work late. So we often play basketball and do our homework together after school. He’s smart and helpful. He often helps me with my math. We’ll graduate soon, and I hope to study in the same high school with him.

三、短文复述及提问

第一节 短文复述

参考范文：

It was Peter’s grandpa’s 70th birthday yesterday. They had a party at their house. The party started at five pm. They invited all their relatives. They enjoyed the dinner together. After the dinner, they played some interesting games. Peter’s grandpa got many gifts. Peter’s uncle bought a new mobile phone for him. Peter’s father gave him a pair of sports shoes. Peter drew a picture for him. All of them had a great time.

录音原文：

Hi, everyone. I’m Peter. It was my grandpa’s 70th birthday yesterday. We had a party at our house. The party started at five pm. We invited all our relatives. My mother prepared a lot of delicious food. We enjoyed the dinner together. After the dinner, we played some interesting games. My grandpa got many gifts. My uncle bought a new mobile phone for him. My father gave him a pair of sports shoes. I drew a picture for my grandpa. All of us had a great time.

第二节 提问

参考答案：

11. What’s your painting about?
12. Do you live with your grandparents?

Appendix 12 English Zhongkao written test 2021

深圳市 2021 年初中毕业生学业水平考试
英语

第一部分 选择题（50 分）

I. 完形填空 （10 分）

阅读下面短文，从短文后所给的 A、B、C、D 四个选项中选出能填入相应空白处的最佳选项，并在答题卡上将相应字母编号涂黑。（共 10 小题，每小题 1 分）

It started 30 years ago with a squirrel. A few months after I had moved into a downtown __1__, my next-door neighbor, Nicole Figaro, knocked, asking for help with a squirrel that had gotten into her flat.

I frightened the unexpected __2__ away and made a new friend. Soon Nicole was inviting me over for dinner or __3__ home-made food at my door. When I went to my beach house, Nicole helped __4__ my mail and water my plants. And when she traveled as a flight attendant, I did the same for her. Wherever she went, she'd __5__ to bring me wonderful gifts, but her smile has been the greatest of all.

In 1997, I had a lovely child and it was Nicole who took care of me and him. Years later, I returned the favour. I helped out when her husband was sick, __6__ her and giving her support. More and more, we drew inspiration from each other.

__7__, the pandemic hit. My husband was out of work and my son couldn't be back to Britain because of the virus. I was __8__ about them. Nicole helped my family to prepare for the pandemic and shared necessities with me. Without her, I couldn't go through __9__ times.

A __10__ might separate our flats and masks can separate the viruses, but nothing is able to separate our hearts.

1. A. house	B. flat	C. street	D. area
2. A. customer	B. host	C. guest	D. enemy
3. A. selling	B. making	C. posting	D. leaving
4. A. look for	B. copy down	C. pick up	D. hold out
5. A. forget	B. keep	C. think	D. remember
6. A. helping out	B. staying with	C. smiling at	D. depending on
7. A. Especially	B. Actually	C. Generally	D. Unluckily
8. A. worried	B. mad	C. nervous	D. serious
9. A. exciting	B. simple	C. difficult	D. good
10. A. room	B. wall	C. yard	D. garden

II. 阅读理解 （40 分）

第一节 阅读下列短文，从每小题所给的 A、B、C、D 四个选项中选出最佳选项，并在答题卡上将相应字母编号涂黑。（共 15 小题，每小题 2 分）

A

Shashiyu was once a poor village in Hebei Province, but now it has changed into a rich and livable place.

In the 1940s, villagers in Shashiyu had little food and few clothes, but they had a strong wish for a better life. Zhang Guishun, the Party secretary of Shashiyu, encouraged his villagers to do their best to pull themselves out of poverty. "In the ancient Chinese story, Yu Gong could move mountains with his strong will and hard work, why can't we? Nobody was born to be poor!" Hearing his words, the villagers decided to make a difference.

Led by the Communist Party of China (中国共产党), the villagers carried water and soil to their village from faraway places to improve their land. From 1966 to 1971, they reclaimed (开垦) lots of land and greatly improved their life.

However, the village encountered a new problem two decades later as the environment became heavily polluted by chemical plants and mine refineries built in 1990s.

The polluting factories were shut down in 2004, and villagers started grape cultivation, which soon became a main industry in Shashiyu.

In 2009, the city-level government invested over 1 million yuan ($152,723) to change the village's exhibition hall into a museum in memory of development efforts made by earlier generations.

After 10 years, the village was called "National Forest Village".

Zhang said that the village started a yearly tourism festival in 2015 to attract travelers to pick grapes. "The grapes could be sold at a better price once Shashiyu becomes famous through this festival. Our villagers can then live even better lives," he said.

11. Why did Zhang Guishun use the story of Yu Gong to encourage the villagers?

A. Because Yu Gong had strong will and hard work.

B. Because Yu Gong was born poor too.

C. Because the villagers worshiped Yu Gong.

D. Because Yu Gong became rich at last.

12. From 1966 to 1971, what did the villagers do to improve their living situation?

A. They planted grapes to earn money.

B. They carried water and soil to their village to improve their lands.

C. They moved to another place to live a better life.

D. They built a lot of factories.

13. Which word can be used to describe the villagers?

A. Hard-working. B. Lazy. C. Kind-hearted. D. Stupid.

14. When was Shashiyu called "National Forest Village"?

A. In 2009. B. In 2004. C. In 2015. D. In 2019.

15. Where can you probably read the passage?

A. In a storybook. B. In a newspaper.

C. In a sports magazine. D. In a science book.

B

When I was a girl growing up, I cannot once ever remember either my mum or my grandma wasting food.

Anything we didn't eat at one meal was leftovers (剩菜). I can remember my grandma making a huge pot of potatoes. We would all eat until we were full enough, but there were always about half of the potatoes left over. A few days later, Grandma would take those potatoes out of the fridge, boil some noodles, add some vegetables and mix them all together. And I also remember when I watched my mum fry bacon (腌培根) for us in the mornings, she would always take the oil carefully and pour it into a bottle. Then she would use it later to add flavor to so many other dishes. I was an adult before I realized that potatoes didn't actually taste like bacon!

I learned their lessons well, and after I grew up, I tried to never waste food myself. I always planned the week's meals ahead of time and only bought what was on my shopping list so that nothing went to waste. Every meal went into our stomachs and any leftovers were later eaten by either myself, my boys, my husband or my dogs. To me, throwing food into the bin was just wrong.

However, when it comes to living, there are no leftovers. Life is just like a feast. Each moment that you don't live is lost forever. Life cannot be saved or stored. Each day is a fresh beginning. Live each moment of your life to the fullest, then.

16. The writer's grandma used the leftover potatoes to _____.

A. mix with noodles　　B. boil some vegetables

C. fry bacon　　D. throw away

17. Why did the potatoes cooked by the writer's mum taste like bacon?

A. Because she mixed the bacon with potatoes.

B. Because she put potatoes into the containers.

C. Because she was good at cooking.

D. Because she poured the bacon oil into potatoes.

18. According to the passage, how did the writer save food in her way?

A. By shopping ahead of time.　　B. By throwing it into the litter bin.

C. By following her mother's shopping list.　　D. By planning the next week's food.

19. What does the underlined word "feast" in the last paragraph probably mean?

A. Sweat dream.　　B. Large meal.

C. Happy story.　　D. Beautiful picture.

20. What can we learn from the last two paragraphs?

A. It is never too old to learn.　　B. Eat to live, but not live to eat.

C. Neither food nor life should be wasted. D. There is no such thing as a free lunch.

C

Can plants talk? Modern research has found something amazing: they do communicate with each other.

It has been known for some time that plants use chemicals to communicate with each other. This happens when a plant gets attacked by insects. The plant gives out chemicals from the leaves that are being eaten. This is like a warning, or a call for help: "I'm being attacked!" When another plant gets the chemicals, it starts to give out its own, different chemicals. Some of these chemicals drive insects away. Others attract the wasps (黄蜂; 蜜蜂)! The wasps kill the insects that are eating the plants. Scientists hope to learn more about this plant warning system, so that we can use it to grow more crops.

More surprisingly, plants also use sound to communicate. People can't hear these sounds, but plants are making them. Some plants make noises with their roots. Corn and chili plants do this. Some trees make clicking noises when there is not enough water.

Most surprisingly of all, plants have an amazing system of communication that can link nearly every plant in a forest. Scientists call this system the "wood wide web". The wood wide web is linked underground by fungi (蘑菇). It links the roots of different plants to each other. It is in some ways similar to the Internet we use. Using the wood wide web, plants can share information and even food with each other. However, it may lead to bad effects. Plants may use it to steal food from each other, or spread chemicals to attack other plants. Perhaps one day scientists will learn how to create a "firewall" to help prevent these attacks within the wood wide web.

Scientists are learning more every day about the secret ways in which plants talk to each other. Maybe one day we will know enough about plant communication to be able to "talk" with them ourselves.

21. What will plants do when they are attacked by insects?

A. They will kill the insects by themselves.

B. They will control the wasps to kill the insects.

C. They will send out signals to ask for help.

D. They will produce chemicals as a warning or a call for help.

22. How are the wood wide web linked by fungi?

A. By making noises with their roots.

B. By connecting the roots of different plants to each other.

C. By sharing food and information with each other.

D. By spreading chemicals to each other.

23. What is the writer's intention of the last paragraph?

A. To give people hope for further studies.

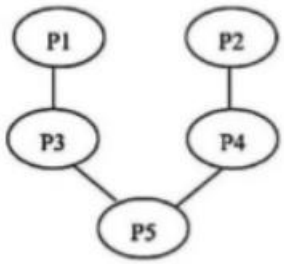
B. To praise scientists for their great achievements.

C. To call on people to protect the plants on Earth.

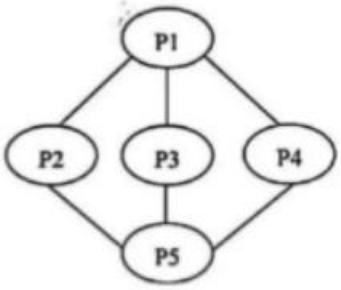
D. To inspire people to communicate with plants.

24. What is the structure of the passage?

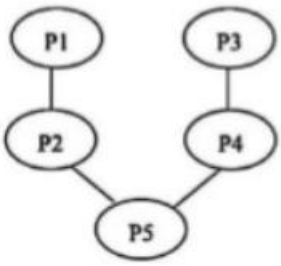
A.



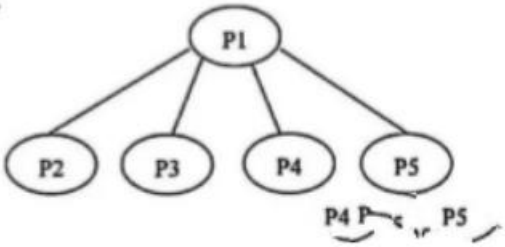
B.



C.



D.



25. Which is the best title for the passage?

A. The Secret Language of Plants

B. The Study about the Plant Warning System

C. The Plants that Give Warnings

D. Why Scientists Do Research about Plants

第二节 短文填空

以下是一则短文，请阅读短文，并从下列方框里的六个句子中选择五个还原到原文中，使原文的意思完整、连贯，并在答题卡上将相应的字母编号涂黑。（共 5 小题，每小题 1 分）

A. Take a backview photo.
B. Not every photo needs to have you stand.
C. This is a fantastic way to increase your beauty.
D. But what if the photos are not as pretty as you expected?
E. But how do you choose and share your photographs properly?
F. For example, you can look into the sky and let the camera catch your side face.

If you have planned a trip during summer holiday, certainly you will want to share some photos on WeChat. 26. _____ Don't worry! Here are some tips you can follow to take wonderful photos.

Turn back and smile.

While on the road, the moment you turn back with a smile is the most impressive. Just walk away from the camera and do a turn-back for the most natural effect. You can also pretend to hear your name called, turn back, and run your fingers through long hair. 27. _____

Look away from the camera.

There are many ways for you to avoid looking at the camera. 28. _____ Looking down into the distance or looking out of the window is a good option. If there is nothing to see, you can also choose to close your eyes and enjoy this moment of quietness.

29. _____

Are you a little shy when facing the camera? Taking a backview photo is a good choice, which can leave room for imagination and bring a different feeling as well.

Take a seat.

30. _____ If there are good places to sit down, please do it! You can pose in a variety of positions with your arms and legs.

Taking the perfect photo takes time and practice. If you follow these simple tricks above, your photos will start coming out beautifully! Do not hesitate to go out with your friends and take more photos.

第三节 信息匹配

阅读下面商品介绍，为五个人选出合适他们的物品，并在答题卡上将相应的字母编号涂黑。（共 5 小题，每小题 1 分）

A.

Future Astronaut
Do you want to be an astronaut in the future? If you do, our club is your best choice!
Here you can explore the universe and the unknown mystery behind it!

B.

Cooking Lesson
Are you good at cooking? Do you enjoy delicious food? Learn to cook in Room 401 every Tuesday. Come here for more information if you are interested in it.

C.

English Corner
Are you an English lover? Come and join our English corner! It's a good place for you to practice spoken English and share experience in learning English.

D.

A Music Video Show
Do you like music? You can have a great time in our music club. We'll make a music video show together!

E.

Basketball Club
Do you like playing basketball? Welcome to join us! We'll invite Yi Jianlian to play a basketball match with us.

F.

Beijing Opera Lesson
Do you want to learn Beijing Opera? Join us at Shenzhen Grand Theater! We'll have teachers here to share professional skills.

( ) 31. Justin is crazy about playing basketball and he is a huge fan of Yi Jianlian.
( ) 32. Alice likes Indian food and Chinese food. She wants to learn to cook wonderful dishes.
( ) 33. Jenny is fond of singing and dancing. She wants to play a role in a music video.
( ) 34. Henry is interested in learning Beijing Opera. He wants to learn from teachers.
( ) 35. Daniel shows great interest in space and hopes he will visit another planet one day.

第二部分 非选择题（25 分）

III. 语法填空（10 分）

阅读下面短文，在空白处填入一个适当的单词或括号内单词的正确形式，并将答案写在答题卡规定的位置上。(共 10 小题，每小题 1 分)

In an old building in Peking University, a yellow light shone into the old and dark hours of the night. The famous translator Xu Yuanchong 36. _____ (sit) behind the window and staring at the computer screen. He enjoyed working by 37. _____ (he) from 10:00 p.m. to 4:00 a.m. For him, the quiet night was a 38. _____ (value) time to focus on translation. He was exploring how 39. _____ (spread) Chinese culture abroad by translating Chinese beauty into Western beauty.

Xu was born into a literary family in Nanchang 40. _____ April 18th, 1921. When he was little, he 41. _____ (influence) by his mother who was interested in literature. Xu began to translate Chinese poetry into both English and 42. _____ (France) when he was still a teenager. He soon learned that Chinese and European languages were very different. Only about half of the words in these languages could be translated word-for-word. This is why translators should have 43. _____ creative mind. They are always trying to improve their translations to make sure they are correct and beautiful.

Xu's motto is: *Good, better, best. Never let it rest. Until your good is better, and your better best.* 44. _____ Xu was widely considered to be the best, he never stopped trying to become better. The more you learn and the 45. _____ (hard) you study, the better you will become. There is always room for improvement.

IV. 书面表达（15 分）

深圳是一座科技之城、创新之城，为培养学生对科技的兴趣，你的学校上周举办了科技节。假定你是你校英语报的小记者，请你编写一篇新闻报道，介绍这次活动。

要点：

1：活动目的

2：活动项目：①模型制作比赛

②……

3：活动亮点：学生发明的服务于校园的机器人（外表，功能）

4：学生感受

提示词： 1. 科技节 science festival 2. 实验 experiment

Shenzhen is well-known as a city of technology and innovation. Last week, ___